

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 14:02:19 ; Search time 3422 seconds
(without alignments)
18755.907 Million cell updates/sec

Title: US-09-818-990B-1
Perfect score: 3963
Sequence: 1 atcgacgacgacgacataga.....tggagagtgatgaacttttaa 3963

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estcom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	478	12.1	553	12	BF126187 601650450
2	291	7.3	931	14	BQ218668 AGENCOURT
3	270	6.8	503	17	AQ621045 HS_2221_B
4	212	5.3	481	9	AA179599 zp49f09.f
5	165	4.2	331	17	AQ167332 HS_3173_A
6	159	4.0	439	17	AQ768080 HS_3097_B

BM968035	LM24HW007
AG083688	Pan trogl
AW918754	EST350058
BQ559749	H4060B02-
AZ889597	RPCI-24-1
BB134094	BB134094
BH261787	CH230-90K
AW315222	12352 MAR
BB633981	BB633981
AZ330563	LM0056C01
AL298368	Tetraodon
AL233081	Tetraodon
AL226160	Tetraodon
AQ072415	HS_3024_A
BM487967	pqm2n.pk0
AQ002636	CIT978SK
BJ070056	BJ070056
BJ069302	BJ069302
AL226161	Tetraodon
BJ065434	BJ065434
BI449569	daa96hl1.
BB196537	BB196537
AG159206	Fan trogl
AA148157	z031c02.f
AQ178082	HS_2217_B
BE938709	KC3-TN009
BH548675	BOGUP89TR
AV665761	AV665761
AQ490686	RPCI-11-2
BF152604	uz35e09.y
AG085351	Pan trogl
BG546584	602574439
BH523182	BOHEJ82TF
BI662601	603302549
BI526367	602925663
BQ895302	AGENCOURT
BQ677523	AGENCOURT
BI691193	603314585
BO684323	AGENCOURT
AK010350	muscu
AZ876084	2M0190K19
AW376351	RC5-CT021
AV881051	AV881051
T31374	EST31281 Hu
AI698879	wc95d03.x
AF315705	AV315705
BF963879	QV2-NN004
BB201416	BB201416
AV343061	AV343061
BB109550	BB109550
D56103	HUM416A01B
BF556746	UI-R-El-f
BH645889	BOHXQ22TR
BB540471	BB540471
BB194413	BB194413
BB240183	BB240183
BF956270	QV2-NN004
BB242103	BB242103
BB401021	BB401021
BB201629	BB201629
BB240329	BB240329
BB243769	BB243769
BB449771	BB449771
BB203035	BB203035
BE300487	ba83d01.x
BB124669	BB124669
AZ609572	LM0434J13
BB234660	BB234660
AI956031	wt34a05.x
BB209138	BB209138
H54549	yq91h05.r1
T33457	EST57947 Hu
AA985151	am75e12.s

c 80	20	0.5	328	9	AA305887	AA305887 EST176886	c 153	20	0.5	432	14	N37003	N37003 yy40a03.s1
c 81	20	0.5	330	9	AA143537	AA143537 zo32b10.s	c 154	20	0.5	434	9	AA171726	AA171726 zo95e01.s
c 82	20	0.5	330	10	AA472766	AA472766 xq19b11.x	c 155	20	0.5	434	10	AW591744	AW591744 xc87g10.x
c 83	20	0.5	331	14	T05375	T05375 EST03264 Fe	c 156	20	0.5	435	14	BM825747	BM825747 k-EST0097
c 84	20	0.5	331	9	AA758826	AA758826 ah76d05.s	c 157	20	0.5	437	9	AA113239	AA113239 zn66d05.s
c 85	20	0.5	331	10	BB550055	BB550055 BB550055	c 158	20	0.5	437	12	BG225506	BG225506 kp68f10.y
c 86	20	0.5	333	9	AA070066	AA070066 zm50d04.r	c 159	20	0.5	437	14	W43020	W43020 zc84e11.s1
c 87	20	0.5	333	9	AA634124	AA634124 ac34e03.s	c 160	20	0.5	437	14	W43024	W43024 zc84f11.s1
c 88	20	0.5	334	10	BB519999	BB519999 BB519999	c 161	20	0.5	438	9	AA399590	AA399590 zt93d04.s
c 89	20	0.5	334	14	W80897	W80897 zd83e06.r1	c 162	20	0.5	439	10	AW275274	AW275274 xv79a12.x
c 90	20	0.5	336	9	AA070835	AA070835 zm60d04.s	c 163	20	0.5	441	14	N20632	N20632 yx40a05.s1
c 91	20	0.5	338	9	AI687776	AI687776 tp03c12.x	c 164	20	0.5	442	9	AA457447	AA457447 aa86d09.r
c 92	20	0.5	341	9	AI581154	AI581154 t194b03.x	c 165	20	0.5	442	9	AA578597	AA578597 nk92h06.s
c 93	20	0.5	342	9	AI268504	AI268504 qo37c06.x	c 166	20	0.5	442	14	W60512	W60512 zc99e01.s1
c 94	20	0.5	342	17	AZ949697	AZ949697 2M0213N03	c 167	20	0.5	443	14	H80208	H80208 yu56h07.r1
c 95	20	0.5	346	14	F10573	F10573 HSC3HA112 n	c 168	20	0.5	443	14	W42769	W42769 zc84g02.s1
c 96	20	0.5	349	14	R68587	R68587 yH9f04.s1	c 169	20	0.5	444	9	AI199980	AI199980 qf84g08.x
c 97	20	0.5	353	14	BM999157	BM999157 UI-H-DI0-	c 170	20	0.5	444	13	BG926835	BG926835 HNC24-1-H
c 98	20	0.5	354	14	N26423	N26423 yx23a06.r1	c 171	20	0.5	444	14	H11623	H11623 ym17a02.s1
c 99	20	0.5	357	9	AI864587	AI864587 wl54f09.x	c 172	20	0.5	445	9	AI423339	AI423339 tf36c06.x
c 100	20	0.5	360	9	AI365397	AI365397 qz08d01.x	c 173	20	0.5	445	14	BM758152	BM758152 k-EST0037
c 101	20	0.5	361	9	AI721077	AI721077 as69g07.x	c 174	20	0.5	447	9	AI214594	AI214594 qm28h03.x
c 102	20	0.5	362	14	N67089	N67089 yz45f04.s1	c 175	20	0.5	447	14	R61550	R61550 ym16g04.s1
c 103	20	0.5	364	12	BG475893	BG475893 mac03f09.	c 176	20	0.5	447	14	W42768	W42768 zc84g01.s1
c 104	20	0.5	366	12	BG054566	BG054566 7f18g04.x	c 177	20	0.5	448	9	AI632835	AI632835 tz32g10.x
c 105	20	0.5	367	14	H13759	H13759 yJ09g01.r1	c 178	20	0.5	449	9	AA412702	AA412702 zu12g07.s
c 106	20	0.5	369	9	AI700275	AI700275 wd05g01.x	c 179	20	0.5	449	9	AA412703	AA412703 zu12g08.s
c 107	20	0.5	369	12	BF896448	BF896448 RC2-WT014	c 180	20	0.5	449	14	N23217	N23217 yx70d12.s1
c 108	20	0.5	370	12	BF896482	BF896482 RC2-WT014	c 181	20	0.5	451	10	AW265015	AW265015 xq59h02.x
c 109	20	0.5	372	9	AI022053	AI022053 ow72d07.x	c 182	20	0.5	454	14	BM761438	BM761438 k-EST0042
c 110	20	0.5	372	10	AW296975	AW296975 UI-H-BW0-	c 183	20	0.5	455	12	BG179037	BG179037 g02330745
c 111	20	0.5	375	12	BF843256	BF843256 PM0-HT107	c 184	20	0.5	456	9	AI129577	AI129577 qc40c08.x
c 112	20	0.5	376	14	W05302	W05302 za83h01.r1	c 185	20	0.5	458	10	AV713167	AV713167 av713167
c 113	20	0.5	377	9	AA857896	AA857896 oe85d02.s	c 186	20	0.5	458	14	H05956	H05956 yl76d02.s1
c 114	20	0.5	378	9	AI152320	AI152320 ud87h11.r	c 187	20	0.5	460	9	AI245353	AI245353 qx28g12.x
c 115	20	0.5	381	10	AW244099	AW244099 xs50h10.x	c 188	20	0.5	461	9	AA725804	AA725804 ai23d08.s
c 116	20	0.5	381	10	AA400249	AA400249 zu2e09.s	c 189	20	0.5	462	10	AW102901	AW102901 xd40a09.x
c 117	20	0.5	381	10	AA467473	AA467473 he19a05.x	c 190	20	0.5	464	9	AI815029	AI815029 wk70b09.x
c 118	20	0.5	383	9	AA770006	AA770006 ah81g06.s	c 191	20	0.5	464	14	H94909	H94909 yu57b07.r1
c 119	20	0.5	383	9	AI025877	AI025877 ow12e10.s	c 192	20	0.5	465	9	AI539659	AI539659 tp60h10.x
c 120	20	0.5	383	14	BM761821	BM761821 k-EST0042	c 193	20	0.5	485	13	BM664112	BM664112 UI-E-CK1-
c 121	20	0.5	384	14	C74987	C74987 C74987 Huma	c 194	20	0.5	485	14	H63616	H63616 yr26g02.r1
c 122	20	0.5	384	14	W02651	W02651 zc64h10.s1	c 195	20	0.5	466	9	AA503401	AA503401 ne45h11.s
c 123	20	0.5	395	9	AA159427	AA159427 zo61g08.r	c 196	20	0.5	467	9	AA405085	AA405085 zu12g07.r
c 124	20	0.5	396	9	AI559943	AI559943 tq77a09.x	c 197	20	0.5	467	14	N28783	N28783 yx69h12.r1
c 125	20	0.5	398	10	AW136170	AW136170 UI-H-B11-	c 198	20	0.5	467	17	AQ126297	AQ126297 HS.3036.B
c 126	20	0.5	400	9	AI843050	AI843050 UI-M-AK1-	c 199	20	0.5	468	14	H03624	H03624 yj37b09.r1
c 127	20	0.5	403	14	BQ219817	BQ219817 AGENCOURT	c 200	20	0.5	469	14	R44291	R44291 y934b12.s1
c 128	20	0.5	404	14	R39327	R39327 yd01f07.s1	c 201	20	0.5	470	9	AA644539	AA644539 ab62g08.s
c 129	20	0.5	406	10	AW572448	AW572448 xq18b11.x	c 202	20	0.5	470	9	AI742544	AI742544 wb55c01.x
c 130	20	0.5	406	12	BF382888	BF382888 601816711	c 203	20	0.5	471	9	AA398996	AA398996 zt93d04.r
c 131	20	0.5	407	12	BF898650	BF898650 QV1-WT022	c 204	20	0.5	471	9	AA558422	AA558422 nk38c01.s
c 132	20	0.5	408	9	AA143786	AA143786 zo31c02.s	c 205	20	0.5	473	9	AA948026	AA948026 oq58h12.s
c 133	20	0.5	408	10	AW238387	AW238387 xp24f03.x	c 206	20	0.5	475	9	AA843760	AA843760 aj18b12.s
c 134	20	0.5	409	9	AA903459	AA903459 ok58g01.s	c 207	20	0.5	475	14	C06009	C06009 C06009 Huma
c 135	20	0.5	411	9	AI267718	AI267718 ag93h06.x	c 208	20	0.5	476	9	AI829320	AI829320 wk58g07.x
c 136	20	0.5	412	9	AA854663	AA854663 aj63g05.s	c 209	20	0.5	479	9	AI820088	AI820088 wq56b09.x
c 137	20	0.5	412	14	H13707	H13707 yJ09g01.s1	c 210	20	0.5	479	9	AA526754	AA526754 nl56a12.s
c 138	20	0.5	414	12	BE762819	BE762819 QV3-WT002	c 211	20	0.5	480	10	BE393233	BE393233 601306438
c 139	20	0.5	414	14	T96399	T96399 ye34c05.s1	c 212	20	0.5	482	10	BE200481	BE200481 ug63e06.x
c 140	20	0.5	417	9	AA732754	AA732754 ai28c10.s	c 213	20	0.5	482	14	C06159	C06159 C06159 Huma
c 141	20	0.5	419	14	H94852	H94852 yu57b07.s1	c 214	20	0.5	483	14	N30039	N30039 yw80c11.s1
c 142	20	0.5	421	9	AA255029	AA255029 mz85b12.r	c 215	20	0.5	484	10	AW009816	AW009816 ws87h03.x
c 143	20	0.5	423	10	AW377536	AW377536 QV0-CT022	c 216	20	0.5	485	9	AI775549	AI775549 EST257049
c 144	20	0.5	424	9	AI335353	AI335353 qo88d06.x	c 217	20	0.5	489	14	BM840929	BM840929 k-EST0118
c 145	20	0.5	424	14	T52636	T52636 ya72a08.s1	c 218	20	0.5	491	9	AA151703	AA151703 z139f04.s
c 146	20	0.5	426	9	AI826788	AI826788 wk56c06.x	c 219	20	0.5	491	14	BM793241	BM793241 k-EST0073
c 147	20	0.5	427	10	AW263394	AW263394 xol5e11.x	c 220	20	0.5	491	14	T32456	T32456 EST48711 Hu
c 148	20	0.5	427	17	BH115735	BH115735 RPC1-24-2	c 221	20	0.5	493	14	BM968955	BM968955 UI-CF-DU1
c 149	20	0.5	429	10	AW470683	AW470683 ha31c11.x	c 222	20	0.5	494	9	AA166828	AA166828 zq39f01.s
c 150	20	0.5	431	9	AA311653	AA311653 EST182513	c 223	20	0.5	494	14	BM841252	BM841252 k-EST0118
c 151	20	0.5	431	10	BE246590	BE246590 TCBAPlE49	c 224	20	0.5	495	9	AA126150	AA126150 z186c05.s
c 152	20	0.5	432	12	BG089773	BG089773 mab80h03.	c 225	20	0.5	499	10	AW513554	AW513554 xo86c11.x

518 19 0.5 518 13 BI085469
519 19 0.5 519 14 HI8557
520 19 0.5 520 17 A0188298
521 19 0.5 521 17 A0625532
522 19 0.5 522 9 A1651864
523 19 0.5 523 13 BI816996
524 19 0.5 524 10 AV566804
525 19 0.5 525 12 BG666831
526 19 0.5 526 17 A0611342
527 19 0.5 527 17 A2827736
528 19 0.5 528 10 A0403729
529 19 0.5 529 13 BI017565
530 19 0.5 530 9 A1249796
531 19 0.5 531 9 A0398950
532 19 0.5 532 12 BF059606
533 19 0.5 533 12 BG672378
534 19 0.5 534 9 A0429894
535 19 0.5 535 10 A0501570
536 19 0.5 536 17 BH281312
537 19 0.5 537 13 BI017483
538 19 0.5 538 12 BF257366
539 19 0.5 539 9 A023911
540 19 0.5 540 12 BF434653
541 19 0.5 541 13 BG977838
542 19 0.5 542 13 BI868014
543 19 0.5 543 17 A0721499
544 19 0.5 544 13 BI016752
545 19 0.5 545 14 BQ336080
546 19 0.5 546 17 A0669972
547 19 0.5 547 10 A0673503
548 19 0.5 548 13 BI085186
549 19 0.5 549 17 A2404731
550 19 0.5 550 12 BF224139
551 19 0.5 551 14 BQ791252
552 19 0.5 552 17 A0622118
553 19 0.5 553 12 BG696554
554 19 0.5 554 17 A2767436
555 19 0.5 555 13 BI020532
556 19 0.5 556 10 AV947010
557 19 0.5 557 13 BI019275
558 19 0.5 558 13 BM554137
559 19 0.5 559 17 A0511922
560 19 0.5 560 10 A0852506
561 19 0.5 561 17 A2744481
562 19 0.5 562 14 BI9673
563 19 0.5 563 12 BG108092
564 19 0.5 564 9 A1595950
565 19 0.5 565 13 BM343810
566 19 0.5 566 12 BF259025
567 19 0.5 567 13 BI016642
568 19 0.5 568 13 BM020892
569 19 0.5 569 13 BI016608
570 19 0.5 570 13 BI085523
571 19 0.5 571 13 BI045438
572 19 0.5 572 17 BH561225
573 19 0.5 573 17 BG799692
574 19 0.5 574 13 BI020223
575 19 0.5 575 12 BG681621
576 19 0.5 576 13 BI241312
577 19 0.5 577 10 A017848
578 19 0.5 578 10 A0180851
579 19 0.5 579 17 A2022407
580 19 0.5 580 13 BI020723
581 19 0.5 581 9 A1042341
582 19 0.5 582 13 BI085313
583 19 0.5 583 9 A1400828
584 19 0.5 584 10 A058699
585 19 0.5 585 13 BI017698
586 19 0.5 586 13 BI045141
587 19 0.5 587 13 BI085119
588 19 0.5 588 13 BI020625
589 19 0.5 589 14 BM795050
590 19 0.5 590 13 BI866952

c 591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663

AA410991 zv03a10.s
A2464702 IM0274008
AG131039 Pan trogl
AZ633587 IM0488J24
BG676137 602622534
BM290756 EST577290
AZ461188 IM0266N22
BH210310 Sml-58L10
BG675764 602621991
AUI33678 AUI33678
BM018305 603645961
AZ447165 IM0243119
AZ452262 IM025H02
AA573374 nk98H06.s
AA349355 AV349355
AV832799 AV832799
BJ483997 BU483997
BQ007582 UI-H-ED1-
BH210319 Sml-58L10
AI911264 wg33e10.x
AJ396327 AJ396327
BM018503 603646331
BG709312 602673326
BM049541 603623482
BH285147 CH230-39J
BM017771 603645194
BM020804 603644454
BI085269 602870308
BM016580 603642909
BM020807 603644457
AL682177 AL682177
BI969792 GM830009A
AZ897857 RPCI-24-2
BG167189 602345324
BM018021 603645563
AZ574057 324PVD05
BG676859 602623461
BM020528 603649257
BM624745 BM624745
BG677813 602625492
BM016807 603643286
BM020740 603644261
BM049506 603623444
BH094186 RPCI-24-2
AUI37100 AUI37100
BG675822 602621861
BU036945 946131A03
BG681956 602629902
AL633398 AL633398
BB197831 BB197831
BM049538 603623479
BG678568 602624248
BM017453 603644558
BM36714 170006875
AV724825 AV724825
AV686595 AV686595
BH866338 LB00726a.
BB239528 BB239528
AG062467 Pan trogl
BG436071 602508858
BH362134 CH230-203
BB629362 BB629362
AW969104 EST381181
BE650294 UI-M-BH2.
BI085083 602870619
BH274863 CH230-701
BG676482 602622941
BG676142 602623539
BG677209 602623903
BG743467 602634986
BE293048 601106037
BQ199777 UI-R-DQ1-
BG678505 602626176

664	19	0.5	691	12	BG680300	19	0.5	737	19	0.5	784	10	BE564158
665	19	0.5	691	12	BG681163	19	0.5	738	19	0.5	784	13	BM018292
666	19	0.5	691	13	BM020092	19	0.5	739	19	0.5	785	13	BI088167
667	19	0.5	691	14	BM15656	19	0.5	740	19	0.5	786	13	BM018298
668	19	0.5	693	12	BG675176	19	0.5	741	19	0.5	789	12	BG678678
669	19	0.5	693	12	BG677278	19	0.5	742	19	0.5	793	12	BG677065
670	19	0.5	693	12	BG679368	19	0.5	743	19	0.5	796	12	BG677347
671	19	0.5	693	13	BM016623	19	0.5	744	19	0.5	799	12	BG681200
c 672	19	0.5	693	17	BH465126	19	0.5	745	19	0.5	804	13	BM016600
c 673	19	0.5	694	13	BH464933	19	0.5	746	19	0.5	803	12	BG680767
c 674	19	0.5	694	13	BH484933	19	0.5	747	19	0.5	804	13	BM019893
c 675	19	0.5	694	13	BM019229	19	0.5	747	19	0.5	805	13	BM016607
c 676	19	0.5	695	9	AI860344	19	0.5	748	19	0.5	806	12	BG680691
c 677	19	0.5	695	12	BG681578	19	0.5	749	19	0.5	806	17	AG139540
c 678	19	0.5	695	13	BM182645	19	0.5	750	19	0.5	806	17	AG139540
c 679	19	0.5	696	12	BG681221	19	0.5	751	19	0.5	806	17	AG139540
c 680	19	0.5	696	12	BM075584	19	0.5	752	19	0.5	807	12	BG679950
680	19	0.5	700	12	BF982856	19	0.5	753	19	0.5	807	13	BI085519
681	19	0.5	700	13	BM019228	19	0.5	754	19	0.5	808	12	BG743087
682	19	0.5	702	13	BM020102	19	0.5	755	19	0.5	808	13	BM018490
c 683	19	0.5	703	10	BE533034	19	0.5	756	19	0.5	809	12	BG678884
c 684	19	0.5	706	9	AI745576	19	0.5	757	19	0.5	813	13	BM017878
c 685	19	0.5	707	9	AI598867	19	0.5	757	19	0.5	814	9	AU128441
c 686	19	0.5	707	10	AV943378	19	0.5	758	19	0.5	815	13	BM020243
c 687	19	0.5	707	13	BM018847	19	0.5	759	19	0.5	817	12	BG680341
688	19	0.5	708	17	AG140954	19	0.5	760	19	0.5	818	12	BG682014
689	19	0.5	712	12	BE705630	19	0.5	761	19	0.5	818	12	BF338453
c 690	19	0.5	713	9	AA887587	19	0.5	762	19	0.5	819	12	BF338453
c 691	19	0.5	713	9	AA887587	19	0.5	763	19	0.5	820	13	BM017512
c 692	19	0.5	713	9	AA887587	19	0.5	764	19	0.5	820	13	BM017512
c 693	19	0.5	713	9	AA887587	19	0.5	765	19	0.5	821	13	BM019298
c 694	19	0.5	713	9	AA887587	19	0.5	766	19	0.5	822	13	BI870926
c 695	19	0.5	713	9	AA887587	19	0.5	767	19	0.5	824	12	BG678164
c 696	19	0.5	713	9	AA887587	19	0.5	768	19	0.5	825	13	BM019637
c 697	19	0.5	713	9	AA887587	19	0.5	769	19	0.5	825	13	BM019637
c 698	19	0.5	713	9	AA887587	19	0.5	770	19	0.5	825	13	BM019637
c 699	19	0.5	713	9	AA887587	19	0.5	771	19	0.5	825	13	BM019637
c 700	19	0.5	713	9	AA887587	19	0.5	772	19	0.5	825	13	BM019637
c 701	19	0.5	713	9	AA887587	19	0.5	773	19	0.5	825	13	BM019637
c 702	19	0.5	713	9	AA887587	19	0.5	774	19	0.5	825	13	BM019637
c 703	19	0.5	713	9	AA887587	19	0.5	775	19	0.5	825	13	BM019637
c 704	19	0.5	713	9	AA887587	19	0.5	776	19	0.5	825	13	BM019637
c 705	19	0.5	713	9	AA887587	19	0.5	777	19	0.5	825	13	BM019637
c 706	19	0.5	713	9	AA887587	19	0.5	778	19	0.5	825	13	BM019637
c 707	19	0.5	713	9	AA887587	19	0.5	779	19	0.5	825	13	BM019637
c 708	19	0.5	713	9	AA887587	19	0.5	780	19	0.5	825	13	BM019637
c 709	19	0.5	713	9	AA887587	19	0.5	781	19	0.5	825	13	BM019637
c 710	19	0.5	713	9	AA887587	19	0.5	782	19	0.5	825	13	BM019637
c 711	19	0.5	713	9	AA887587	19	0.5	783	19	0.5	825	13	BM019637
c 712	19	0.5	713	9	AA887587	19	0.5	784	19	0.5	825	13	BM019637
c 713	19	0.5	713	9	AA887587	19	0.5	785	19	0.5	825	13	BM019637
c 714	19	0.5	713	9	AA887587	19	0.5	786	19	0.5	825	13	BM019637
c 715	19	0.5	713	9	AA887587	19	0.5	787	19	0.5	825	13	BM019637
c 716	19	0.5	713	9	AA887587	19	0.5	788	19	0.5	825	13	BM019637
c 717	19	0.5	713	9	AA887587	19	0.5	789	19	0.5	825	13	BM019637
c 718	19	0.5	713	9	AA887587	19	0.5	790	19	0.5	825	13	BM019637
c 719	19	0.5	713	9	AA887587	19	0.5	791	19	0.5	825	13	BM019637
c 720	19	0.5	713	9	AA887587	19	0.5	792	19	0.5	825	13	BM019637
c 721	19	0.5	713	9	AA887587	19	0.5	793	19	0.5	825	13	BM019637
c 722	19	0.5	713	9	AA887587	19	0.5	794	19	0.5	825	13	BM019637
c 723	19	0.5	713	9	AA887587	19	0.5	795	19	0.5	825	13	BM019637
c 724	19	0.5	713	9	AA887587	19	0.5	796	19	0.5	825	13	BM019637
c 725	19	0.5	713	9	AA887587	19	0.5	797	19	0.5	825	13	BM019637
c 726	19	0.5	713	9	AA887587	19	0.5	798	19	0.5	825	13	BM019637
c 727	19	0.5	713	9	AA887587	19	0.5	799	19	0.5	825	13	BM019637
c 728	19	0.5	713	9	AA887587	19	0.5	800	19	0.5	825	13	BM019637
c 729	19	0.5	713	9	AA887587	19	0.5	801	19	0.5	825	13	BM019637
c 730	19	0.5	713	9	AA887587	19	0.5	802	19	0.5	825	13	BM019637
c 731	19	0.5	713	9	AA887587	19	0.5	803	19	0.5	825	13	BM019637
c 732	19	0.5	713	9	AA887587	19	0.5	804	19	0.5	825	13	BM019637
c 733	19	0.5	713	9	AA887587	19	0.5	805	19	0.5	825	13	BM019637
c 734	19	0.5	713	9	AA887587	19	0.5	806	19	0.5	825	13	BM019637
c 735	19	0.5	713	9	AA887587	19	0.5	807	19	0.5	825	13	BM019637
c 736	19	0.5	713	9	AA887587	19	0.5	808	19	0.5	825	13	BM019637

956	19	0.5	1171	13	BI085315	602870373
957	19	0.5	1174	13	BM549134	AGENCOURT
958	19	0.5	1176	13	BM045270	603623233
959	19	0.5	1176	14	BM911261	AGENCOURT
960	19	0.5	1180	13	BI085106	602870647
961	19	0.5	1190	13	BM549130	AGENCOURT
962	19	0.5	1198	14	BM908893	AGENCOURT
963	19	0.5	1204	13	BM549251	AGENCOURT
964	19	0.5	1204	14	BM914218	AGENCOURT
965	19	0.5	1205	13	BM017689	603645028
966	19	0.5	1207	14	BM910307	AGENCOURT
967	19	0.5	1208	14	BM908519	AGENCOURT
968	19	0.5	1213	14	BM909247	AGENCOURT
969	19	0.5	1222	14	BM908269	AGENCOURT
970	19	0.5	1222	14	BM910574	AGENCOURT
971	19	0.5	1224	14	BM908972	AGENCOURT
972	19	0.5	1234	14	BM911085	AGENCOURT
973	19	0.5	1235	13	BI085338	602870004
974	19	0.5	1242	14	BM913677	AGENCOURT
975	19	0.5	1243	13	BM018595	603646459
976	19	0.5	1243	14	BM911264	AGENCOURT
977	19	0.5	1248	13	BM017879	603645358
978	19	0.5	1251	13	BM045264	603623223
979	19	0.5	1251	14	BM911250	AGENCOURT
980	19	0.5	1252	14	BM908638	AGENCOURT
981	19	0.5	1256	13	BM045182	603623103
982	19	0.5	1260	13	BM549492	AGENCOURT
983	19	0.5	1266	14	BM913626	AGENCOURT
984	19	0.5	1271	13	BM018130	603645718
985	19	0.5	1281	14	BM910859	AGENCOURT
986	19	0.5	1287	14	BM908666	AGENCOURT
987	19	0.5	1289	14	BM910149	AGENCOURT
988	19	0.5	1298	14	BM908291	AGENCOURT
989	19	0.5	1309	14	BM911265	AGENCOURT
990	19	0.5	1316	14	BM913704	AGENCOURT
991	19	0.5	1321	14	BM914078	AGENCOURT
992	19	0.5	1341	13	BI081626	602880271
993	19	0.5	1342	14	BM910331	AGENCOURT
994	19	0.5	1343	14	BM909251	AGENCOURT
995	19	0.5	1351	14	BM908633	AGENCOURT
996	19	0.5	1359	13	BM557819	AGENCOURT
997	19	0.5	1362	14	BM910954	AGENCOURT
998	19	0.5	1377	14	BM909771	AGENCOURT
999	19	0.5	1377	14	BM908569	AGENCOURT
1000	19	0.5	1384	14	BM910087	AGENCOURT

ALIGNMENTS

RESULT 1
BF126187 553 bp mRNA linear EST 24-OCT-2000
LOCUS 601650450F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:3934045 5',
DEFINITION mRNA sequence.
ACCESSION BF126187
VERSION BF126187.1 GI:10965227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 553)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM774 row: d column: 14
High quality sequence stop: 542.
FEATURES
source
1..553
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3934045"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgctcgcc); Site_2: SfiI (ggccattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGGCGGAGGCGGCAGATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 143 a 138 c 153 g 119 t
ORIGIN

Query Match 12.1%; Score 478; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.9e-237;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3486 TGTAGTACCCAAAGAGGTGAAGAACGACCTGTGATCCTGGAGAACTACAGAACTCGGG 3545
Db 1 TGTAGTACCCAAAGAGGTGAAGAACGACCTGTGATCCTGGAGAACTACAGAACTCGGG 60
Qy 3546 TGTTCGCCAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCCGCCACCTGT 3605
Db 61 TGTTCGCCAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCCGCCACCTGT 120
Qy 3606 GTTCTACTGGAGAAAGACAATGAGACCATCCCTTGACACGAGAGAGGATCAGTATGCA 3665
Db 121 GTTCTACTGGAGAAAGACAATGAGACCATCCCTTGACACGAGAGAGGATCAGTATGCA 180
Qy 3666 CCAGGACACACAGGGGTATGCTGCTTCATTTCAGCCAGCCCAAGAAATCAGACGCTGG 3725
Db 181 CCAGGACACACAGGGGTATGCTGCTTCATTTCAGCCAGCCCAAGAAATCAGACGCTGG 240
Qy 3726 ATGCTACACGTTGTTCAGCCAAAGATGAGCGGATCGTGTCTGCTGCACTGCCAGCTGGA 3785
Db 241 ATGCTACACGTTGTTCAGCCAAAGATGAGCGGATCGTGTCTGCTGCACTGCCAGCTGGA 300
Qy 3786 TATATACGCTCAGTGGCACCACATCAGATCCCAACCCCATGTCTGTCCGGCCAGTGGCAG 3845
Db 301 TATATACGCTCAGTGGCACCACATCAGATCCCAACCCCATGTCTGTCCGGCCAGTGGCAG 360
Qy 3846 TCGTACGGATCTCTACCCAGTAAGGACTTGACATATTTTCTGCTTTTCTCCATGGA 3905
Db 361 TCGTACGGATCTCTACCCAGTAAGGACTTGACATATTTTCTGCTTTTCTCCATGGA 420
Qy 3906 AAGCAGCATGGTGTATTCATCTCTTCGAGTGTCTAGTGAGAGTGTGAACATTAA 3963
Db 421 AAGCAGCATGGTGTATTCATCTCTTCGAGTGTCTAGTGAGAGTGTGAACATTAA 478

RESULT 2
BO218668
LOCUS BO218668 931 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7263964 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6017402
5', mRNA sequence.
ACCESSION BO218668
VERSION BO218668.1 GI:204000068
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 931)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13216 row: k column: 03
High quality sequence stop: 501.
FEATURES
source
1..931
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6017402"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies."
BASE COUNT 250 a 225 c 236 g 220 t
ORIGIN
Query Match 7.3%; Score 291; DB 14; Length 931;
Best Local Similarity 100.0%; Pred. No. 8.4e-140;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3673 ACAACAGGTATGCTTCTTCATTCAGCCAGCCAGCAAGAAATCAGACGCTGGATGGTAC 3732
Db 1 ACAACAGGTATGCTTCTTCATTCAGCCAGCCAGCAAGAAATCAGACGCTGGATGGTAC 60
QY 3733 ACCTTGTCAGCCAAAGATGAAGCCGCATCGTGTGTCGACTGCCAGCTGGGATATATAC 3792
Db 61 ACCTTGTCAGCCAAAGATGAAGCCGCATCGTGTGTCGACTGCCAGCTGGGATATATAC 120
QY 3793 GCTCAGTGGCACCATCAGATGCCACGCCCATCTGTCTCGGCCCCAGTGCGCAGTCGCTAC 3852
Db 121 GCTCAGTGGCACCATCAGATGCCACGCCCATCTGTCTCGGCCCCAGTGCGCAGTCGCTAC 180
QY 3853 GGATCTCTACCACTAAGAGACTTGACATATTTCTGCGCTTTCTCCATGGAAAGCAGC 3912
Db 181 GGATCTCTACCACTAAGAGACTTGACATATTTCTGCGCTTTCTCCATGGAAAGCAGC 240
QY 3913 ATGGTGATTCATGCTTCTCGAGTGTAGTGAGAGTGAATGAATTTAA 3963
Db 241 ATGGTGATTCATGCTTCTCGAGTGTAGTGAGAGTGAATGAATTTAA 291
RESULT 3
LOCUS AO621045 503 bp DNA linear GSS 16-JUN-1999
DEFINITION HS_2221_B1_B06_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2221 Col=11 Row=D, DNA sequence.
ACCESSION AO621045
VERSION AO621045.1 GI:5083437
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 503)
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Koller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end web Server: <http://www.htsc.washington.edu>
Plate: 2221 row: D column: 11
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 503.
FEATURES
source
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=2221 Col=11 Row=D"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 154 a 122 c 106 g 115 t 6 others
ORIGIN
Query Match 6.8%; Score 270; DB 17; Length 503;
Best Local Similarity 100.0%; Pred. No. 6.8e-129;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 68 AAACCCAGACATCGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCCAAACCTTGCC 127
Db 93 AAACCCAGACATCGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCCAAACCTTGCC 152
QY 128 ATTTGCGCAGTCTTCTGCGGCGCTGAAGAGGCGGAGGCGCAAGATGACCTTCCAGATC 187
Db 153 ATTTGCGCAGTCTTCTGCGGCGCTGAAGAGGCGGAGGCGCAAGATGACCTTCCAGATC 212
QY 188 TTTTCAGCCTTTTCAGCCCAAGAAGAAATAGACGAAAGTGTCAATTTGGCAGACTGGCCA 247
Db 213 TTTTCAGCCTTTTCAGCCCAAGAAGAAATAGACGAAAGTGTCAATTTGGCAGACTGGCCA 272
QY 248 TCAATTACGACCCCTTTGGAGAAGGCAGATGAACCTCAAGCTAGAAAACGACTTTCTCTG 307
Db 273 TCAATTACGACCCCTTTGGAGAAGGCAGATGAACCTCAAGCTAGAAAACGACTTTCTCTG 332
QY 308 ATCAGATGAACACTCACCCTAAATTTAAAGTT 337
Db 333 ATCAGATGAACACTCACCCTAAATTTAAAGTT 362
RESULT 4
LOCUS AA179599 481 bp mRNA linear EST 31-DEC-1996
DEFINITION ZP49F09.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:612809 5' similar to TF:G1160355 G1160355 UNC-89.; mRNA sequence.
ACCESSION AA179599
VERSION AA179599.1 GI:1760985
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 481)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, J.B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
TITLE Generation and analysis of 280,000 human expressed sequence tags

1026

JOURNAL
MEDLINE
COMMENT

Genome Res. 6 (9), 807-828 (1996)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28kl3 rev2 from Amersham
High quality sequence stop: 417.
Location/Qualifiers
1. .481
/organism="Homo sapiens"
/db_xref="GDB:4644648"
/db_xref="taxon:9606"
/clone="IMAGE:612809"
/clone_lib="Stratagene HeLa cell s3 937216"
/sex="female"
/dev_stage="HeLa S3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR
Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'
2 others

BASE COUNT 129 a 132 c 125 g 93 t
ORIGIN

Query Match 5.3%; Score 212; DB 9; Length 481;
Best Local Similarity 99.7%; Pred. No. 1.2e-98;
Matches 332; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3423 AGGACCTATAAGTGCATCGCTACCAAAACCGGCGAGAAATTCCTTTTGTCTGGAGCT 3482
|||||
Db 148 AGGACCTATAAGTGCATCGCTACCAAAACCGGCGAGAAATTCCTTTTGTCTGGAGCT 207
QY 3483 CTCGTGTAGTACCAAGAGGTGAAGAAGCACCTGTCTCTGGAGAACTACAGAACTG 3542
|||||
Db 208 CTCGTGTAGTACCAAGAGGTGAAGAAGCACCTGTCTCTGGAGAACTACAGAACTG 267
QY 3543 CGGTGTTCCGAAAGCCACCCCGTGGAGCTGGAGTGGCGGTGATAGGCATGCCCCACCC 3602
|||||
Db 268 CGGTGTTCCGAA -GCCACCCCGTGGAGCTGGAGTGGCGGTGATAGGCATGCCCCACCC 326
QY 3603 TGTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGACACGAGAGGATCAGTAT 3652
Db 327 TGTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGACACGAGAGGATCAGTAT 386
QY 3663 GCACGAGGACACACAGGGTATGCTGCTTCTCATTGAGCGACCAAGAAATCAGACGC 3722
Db 387 GCACGAGGACACACAGGGTATGCTGCTTCTCATTGAGCGACCAAGAAATCAGACGC 446
QY 3723 TGGATGTACACGTTGTGACGCAAGATGAAGC 3755
|||||
Db 447 TGGATGTACACGTTGTGACGCAAGATGAAGC 479

RESULT 5
AQ167332/c
LOCUS
DEFINITION
HS_3173_A1_A03_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-3173 Col-5 Row-A, DNA sequence.

ACCESSION AQ167332
VERSION AQ167332.1 GI:3565497
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 331)

1020

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3173 row: A column: 5
Class: BAC ends
High quality sequence stop: 331.
Location/Qualifiers
1. .331
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3173 Col-5 Row-A"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
E-Coli DH10B" 81 a 77 c 86 g 87 t
BASE COUNT
ORIGIN

Query Match 4.2%; Score 165; DB 17; Length 331;
Best Local Similarity 100.0%; Pred. No. 3.9e-74;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3297 ACCGCCCGGAGTGACATGGCTACTCAATGGCAACCTGTGTACAGATGCTCCCA 3356
|||||
Db 165 ACCGCCCGGAGTGACATGGCTACTCAATGGCAACCTGTGTACAGATGCTCCCA 106
QY 3357 CAAGATCTGTGTAGGAGACCGGAGTCCACTCTCTCATTCACCCACTCACTCAGC 3416
|||||
Db 105 CAAGATCTGTGTAGGAGACCGGAGTCCACTCTCTCATTCACCCACTCACTCAGC 46
QY 3417 CGAGCGAGGACCTTAAATGTCATCGCTACCAACAAACCGGCA 3461
|||||
Db 45 CGAGCGAGGACCTTAAATGTCATCGCTACCAACAAACCGGCA 1

RESULT 6
AQ768080
LOCUS
DEFINITION
HS_3097_B1_E02_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-3097 Col-3 Row-J, DNA sequence.

ACCESSION AQ768080
VERSION AQ768080.1 GI:5646196
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 439)

REFERENCE
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3097 row: J column: 3
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 439.
Location/Qualifiers
1. .439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Place=3097 Col=3 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

FEATURES

source

BASE COUNT 105 a 127 c 89 g 117 t 1 others
ORIGIN
Query Match 4.0%; Score 159; DB 17; Length 439;
Best Local Similarity 99.5%; Pred. No. 5.6e-71;
Matches 209; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3284 AGGTGAGTGTTCACGCCGCCGAGCTGACATGGCTACTCAATGGCCAACTGTGTCTAC 3343

Db 67 AGGTGAGTGTTCACGCCGCCGAGCTGACATGGCTACTCAATGGCCAACTGTGTCTAC 126

QY 3344 CAGATGCTCCACAGATGCTGTGAGGAGACCGAGTCCACTCTCTCTCATTTGACC 3403

Db 127 CAGATGCTCCACAGATGCTGTGAGGAGACCGAGTCCACTCTCTCTCATTTGACC 186

QY 3404 CACTCACTACGCGCAGCGAGGACCTATTAAGTGCATCGCTACCAACAAACCGGCAGA 3463

Db 187 CACTCACTACGCGCAGCGAGGACCTATTAAGTGCATCGCTACCAACAAACCGGCAGA 246

QY 3464 ATTCTTTTACTGTGGAGCTCTCTGTAGTAG 3493

Db 247 ATTCTTTTACTGTGGAGCTCTCTGTAGTAG 276

RESULT 7

BM968035

LOCUS

DEFINITION LM24HW00787 Bos taurus LM-24-HW cDNA library Bos taurus cDNA clone
LM-24-HW-007-87 (5'), mRNA sequence.

ACCESSION

BM968035

VERSION

BM968035.1

KEYWORDS

EST.

SOURCE

cow.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 614)

Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,

Sun,S.S. and Cheong,I.C.

Gene Expression Profiling of the Bovine skeletal muscle

Unpublished (2002)

Contact: Dr. Du-Hak Yoon

National Livestock Research Institute, RDA

564 Omockchun-dong, Suwon, 441-350, Korea

Tel: 82 31 290 1593

Fax: 82 31 290 1792

Email: dhyoon@rda.go.kr

Insert Length: 614 Std Error: 0.00

Seq primer: CAGGAACAGCTATGAC

POLYA=NO.

Location/Qualifiers

1. .614

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone="LM-24-HW-007-87 (5')"

/clone_lib="Bos taurus LM-24-HW cDNA library"

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

55; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 3350

CCTCCACAAAGATGCTGTCAGGAGACCGAGTCCACTCTCTCTCATTTGACC 3404

Db 144

CCTCCACAAAGATGCTGTCAGGAGACCGAGTCCACTCTCTCTCATTTGACC 198

RESULT 8

AG083688/c

LOCUS

DEFINITION

Pan troglodytes DNA, clone: PTB-081D08.R, genomic survey sequence.

ACCESSION

AG083688

VERSION

AG083688.1

KEYWORDS

GSS.

SOURCE

Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male

BAC Library clone:PTB-081D08.R.

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 633)

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

[E-mail: chuhpbes@sc.riken.go.jp, URL: http://hgpc.riken.go.jp/.

Tel: 81-45-503-9111, Fax: 81-45-503-9170]

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1. .633

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-081D08.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

169 a 140 c 109 g 213 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

48; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 1077

AGGGTTTCTTCTTCTGACTCAGAGCGGACCCCTTACAAAGGAAGAT 1124

Db 375

AGGGTTTCTTCTTCTGACTCAGAGCGGACCCCTTACAAAGGAAGAT 328

```
RESULT 9
AW918754          376 bp  mRNA  linear  EST 25-MAY-2000
LOCUS
DEFINITION
Rattus norvegicus Rat gene index, normalized rat, norvegicus, Bento Soares
AW918754
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 376)
AUTHORS
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
TITLE
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL
COMMENT
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.
Location/Qualifiers
1..376
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RG1EX58"
/clone_lib="Rat gene index, normalized rat, norvegicus,
Bento Soares"
/tissue_type="mix - brain, ovary, placenta, kidney, lung,
liver, embryo, heart, muscle, spleen"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Estimated insert size approx.1 kb"
BASE COUNT 107 a 96 c 97 g 76 t
ORIGIN
Query Match 1.0%; Score 41; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2890 TTCACCTGCAAAATGTTGGGATACCTGTTCCAAAGGTTTA 2930
|||||
Db 182 TTCACCTGCAAAATGTTGGGATACCTGTTCCAAAGGTTTA 222

RESULT 10
BQ559749          413 bp  mRNA  linear  EST 20-JUN-2002
LOCUS
DEFINITION
H4060B02-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
H4060B02 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 413)
AUTHORS
VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin
, P.R., Stagg, C.A., Bassey, U., Alba, K., Hamatani, T., Kargul, G.J.,
Luo, A.G. and KO, M.S.H.
TITLE
Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set
JOURNAL
COMMENT
Unpublished (2002)
Other_ESTs: H4060B02-3
```

```

Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cdna/NIA\_7\_4k.html for details.
Plate: H4060 row: B column: 02
Seq primer: -21M13 Reverse
High quality sequence stop: 413
POLYA-No.
Location/Qualifiers
1..413
/organism="Mus musculus"
/db_xref="niaEST:H4060B02-5"
/db_xref="taxon:10090"
/clone="H4060B02"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."
BASE COUNT 102 a 123 c 114 g 74 t
ORIGIN
Query Match 1.0%; Score 38; DB 14; Length 413;
Best Local Similarity 100.0%; Pred. No. 7.7e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3196 CTACAGGAACGCTTTTCCGACCACATTTCTCGAGGC 3233
|||||
Db 73 CTACAGGAACGCTTTTCCGACCACATTTCTCGAGGC 110

RESULT 11
AZ889597/c
LOCUS
DEFINITION
RPCI-24-167B6.TJ RPCI-24 Mus musculus genomic clone RPCI-24-167B6,
DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 503)
AUTHORS
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
TITLE
Mouse BAC End Sequences from Library RPCI-24
JOURNAL
COMMENT
Other_GSSs: RPCI-24-167B6.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 167 row: B column: 6
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..503
FEATURES
source
```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPI-24-167B6"
/clone_lib="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPI-24 Mouse BAC Library produced by Pieter de Jong, The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 109 a 103 c 168 g 123 t
ORIGIN
Query Match 0.8%; Score 32; DB 17; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1978 TCCACTCAGTTACACAGCTTCATACCAAGT 2009
|||||
Db 329 TCCACTCAGTTACACAGCTTCATACCAAGT 298

RESULT 12
BB134094
LOCUS
DEFINITION BB134094 513 bp mRNA linear EST 18-OCT-2001
CDNA clone 9830112J03 3', mRNA sequence.
ACCESSION BB134094.2 GI:16266577
VERSION BB134094.2
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 513)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Kouda
Okazaki,Y., Okido,T., Saito,K., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jun 28, 2000 this sequence version replaced gi:8788484.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

```

```

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES
Location/Qualifiers
1..513
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="9830112J03"
/clone_lib="RIKEN full-length enriched, adult male bone"
/sex="male"
/tissue_type="bone"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCCAAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot - 10.0 and subtraction to Rot - 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATTAATATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT 129 a 166 c 113 g 105 t
ORIGIN
Query Match 0.8%; Score 32; DB 10; Length 513;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1978 TCCACTCAGTTACACAGCTTCATACCAAGT 2009
|||||
Db 223 TCCACTCAGTTACACAGCTTCATACCAAGT 254

RESULT 13
BH261787/c
LOCUS
DEFINITION BH261787 521 bp DNA linear GSS 30-NOV-2001
CH230-90K20-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-90K20, DNA sequence.
ACCESSION BH261787
VERSION BH261787.1 GI:17167683
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 521)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 ECORI segment
Unpublished (1999)
Other_GSSs: CH230-90K20.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics

```

The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (<http://www.chori.org/bacpac/orering/information.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 90 row: K column: 20
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1. .521
 /organism="Rattus norvegicus"
 /strain="BN/SSNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-90K20"
 /clone_lib="CHORI-230 Segment 1"
 /sex="Female"
 /cell_type="Brain"
 /note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
 CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
 Pieter de Jong"
 BASE COUNT 145 a 117 c 117 g 142 t
 ORIGIN

Query Match 0.8%; Score 31; DB 17; Length 521;
 Best Local Similarity 100.0%; Pred. No. 0.00036;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3737 TGTGAGCCAGCAATGAAGCGGCATCGTGC 3767
 |||||
 Db 282 TGTGAGCCAGCAATGAAGCGGCATCGTGC 252

RESULT 14
 AW315222 438 bp mRNA linear EST 25-APR-2001
 LOCUS
 DEFINITION 13252 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AW315222
 VERSION AW315222.1 GI:6744478
 KEYWORDS EST.
 SOURCE
 ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 438)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahnenkrug,S.C., Bennett
 ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.

TITLE
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL
 MEDLINE
 COMMENT
 21180013
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 20
 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCCGAGTCACGCG

Plate: 7 row: B column: 24

Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1. .438
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH108"
 /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."
 BASE COUNT 121 a 110 c 119 g 88 t
 ORIGIN

Query Match 0.7%; Score 29; DB 10; Length 438;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3553 GAAGGCCACCCGTCGAGACTGGAGTGCGC 3581
 |||||
 Db 145 GAAGGCCACCCGTCGAGACTGGAGTGCGC 173

RESULT 15
 BB633981 652 bp mRNA linear EST 26-OCT-2001
 LOCUS
 DEFINITION BB633981 RIKEN full-length enriched, adult male spinal cord Mus
 musculus cDNA clone A330094N09 5', mRNA sequence.
 ACCESSION BB633981
 VERSION BB633981.1 GI:16470313
 KEYWORDS EST.
 SOURCE
 ORGANISM

house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 652)
 Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
 ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
 ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)
 TITLE
 JOURNAL
 COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resesgsc.riken.go.jp,
 URL:<http://genome.gsc.riken.go.jp/>
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
 ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 ,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa

K., K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome*. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES

Location/Qualifiers

source

```

1. .652
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A330094N09"
/clone_lib="RIKEN full-length enriched, adult male spinal
cord"
/sex="male"
/tissue_type="spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site1: Sali; Site2: BamHI: cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCATCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length b-
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCCAGTAAATTAATATCCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

```

BASE COUNT	141 a	159 c	143 g	207 t	2 others
ORIGIN					

Query Match 0.7%; Score 29; DB 10; Length 652;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 29; Conservative 0; Mismatches 0; Indels

Qy 3301 CCCCCGGAGCTGACATGGCTACTCAATGG 3329
 |||||
 Db 593 CCCCCGGAGCTGACATGGCTACTCAATGG 621

Search completed: November 30, 2002, 17:03:41
Job time : 3498 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 30, 2002, 19:02:20 ; Search time 36 Seconds
(without alignments)
7555.061 Million cell updates/sec

Title: US-09-818-990B-2
Perfect score: 6890
Sequence: 1 MQDDIEASTISQLLRESY.....MESTMYSCSRVVSDEL 1320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriaph.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6869	99.7	1320	4 Q96KF5	Q96KF5 homo sapien
2	2560	37.2	507	4 Q96K90	Q96K90 homo sapien
3	2213.5	32.1	1106	4 Q8WX93	Q8WX93 homo sapien
4	1666	24.2	772	4 Q9Y2J6	Q9Y2J6 homo sapien
5	1460	21.2	492	11 Q9ET54	Q9ET54 mus musculus
6	1289.5	18.7	385	4 Q9UQF5	Q9UQF5 homo sapien
7	1289.5	18.7	404	4 Q9Y3E9	Q9Y3E9 homo sapien
8	1271.5	18.5	379	11 Q9CWM1	Q9CWM1 mus musculus
9	871.5	12.6	4162	13 Q98918	Q98918 gallus gall
10	826	12.0	2164	13 Q9IAR9	Q9IAR9 gallus gall
11	801.5	11.6	26926	4 Q10466	Q10466 homo sapien
12	801.5	11.6	26926	4 Q8WZB3	Q8WZB3 homo sapien
13	793	11.5	2000	6 Q97791	Q97791 oryctolagus
14	780.5	11.3	5604	4 Q8WZ53	Q8WZ53 homo sapien
15	780.5	11.3	34350	4 Q8WZ42	Q8WZ42 homo sapien
16	729	10.6	498	4 Q9UBF9	Q9UBF9 homo sapien

17	704	10.2	496	11 Q9JIF9	Q9JIF9 mus musculus
18	602.5	8.7	4824	5 Q95YM1	Q95YM1 procambarius
19	602.5	8.7	17352	5 Q95YM2	Q95YM2 procambarius
20	600	8.7	6632	5 Q01761	Q01761 caenorhabdi
21	598	8.7	6632	5 Q17362	Q17362 caenorhabdi
22	585.5	8.5	4816	5 Q8T103	Q8T103 bombyx mori
23	568.5	8.3	1561	11 Q924D2	Q924D2 mus musculus
24	545.5	7.9	4796	5 Q9W055	Q9W055 drosophila
25	545.5	7.9	16215	5 Q9NFS3	Q9NFS3 drosophila
26	542	7.9	4001	5 Q9N2P7	Q9N2P7 drosophila
27	542	7.9	4796	5 Q9NLS8	Q9NLS8 drosophila
28	539	7.8	1802	6 Q28633	Q28633 oryctolagus
29	536	7.8	2673	4 Q96SC3	Q96SC3 homo sapien
30	534	7.8	5636	4 Q96RW7	Q96RW7 homo sapien
31	531.5	7.7	7962	4 Q10465	Q10465 homo sapien
32	503	7.3	2154	4 Q8WZ51	Q8WZ51 homo sapien
33	502	7.3	2341	5 Q19663	Q19663 caenorhabdi
34	501.5	7.3	4203	5 Q965G2	Q965G2 caenorhabdi
35	501.5	7.3	4219	5 Q9NLS8	Q9NLS8 caenorhabdi
36	493.5	7.2	4488	5 Q9TXK2	Q9TXK2 caenorhabdi
37	485	7.0	6620	4 Q96AA2	Q96AA2 homo sapien
38	482	7.0	4650	4 Q15598	Q15598 homo sapien
39	472.5	6.9	3262	11 Q9EQJ5	Q9EQJ5 mus musculus
40	472	6.9	6815	5 Q9I7U4	Q9I7U4 drosophila
41	468.5	6.8	5198	5 Q76518	Q76518 caenorhabdi
42	467	6.8	1232	4 Q8TCG8	Q8TCG8 homo sapien
43	459.5	6.7	1019	4 Q9Y6L9	Q9Y6L9 homo sapien
44	453.5	6.6	1252	4 Q96DN3	Q96DN3 homo sapien
45	435.5	6.3	3354	5 Q8T101	Q8T101 bombyx mori

ALIGNMENTS

RESULT 1

Q96KF5 PRELIMINARY; PRT; 1320 AA.

AC Q96KF5; (TREMREL. 19, Created)
DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
DE Myopalladin.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=SKELETAL MUSCLE;

RX MEDLINE=21206024; PubMed=11309420;

RA Bang M.L., Mudry R.E., McElhinny A.S., Trombitas K., Geach A.J.,

Yamasaki R., Sorimachi H., Granzier H., Gregorio C.C., Labat S.;

RT "Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple

roles in z-disc and I-band protein assemblies.";

RL J. Cell Biol. 153:413-428(2001).

DR EMBL: AF328296; AAK50625.1; -

DR InterPro: IPR000886; ER_target.

DR InterPro: IPR003006; Ig_MHC.

DR Pfam: PF00047; Ig; 5.

DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.

SQ SEQUENCE 1320 AA; 145379 MW; A6579FB164D33B6E CRC64;

Query Match 99.7%; Score 6869; DB 4; Length 1320;

Best Local Similarity 99.5%; Pred. No. 0;

Matches 1314; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MQDDIEASTISQLLRESYLAETRRHGNERSRAEPSSNPFHSPGSGAAEGGGDDL 60

Db 1 MQDDIEASTISQLLRESYLAETRRHGNERSRAEPSSNPFHSPGSGAAEGGGDDL 60

QY 61 PDLSAFLSQEELDESVNLRLAINYDPLEKADETQARKRLSPDQMKHSPNLSFEPNFCQD 120

Db 61 PDLSAFLSQEELDESVNLRLAINYDPLEKADETQARKRLSPDQMKHSPNLSFEPNFCQD 120

Function on human myopalladin


```
|||||
Db 421 NYLQGLDGRPIIAAPVETKMLNLSASEGOLVVFECRVKAGSPKVEWYREGTLIEDSPD 480
QY 481 FRILQKKPRSMARP 494
Db 481 FRILQKKPRSMARP 494
|||||

RESULT 3
Q8WX93
ID Q8WX93 PRELIMINARY; PRT; 1106 AA.
AC Q8WX93;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Myoneurin.
GN MYN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lockwood S.K.; Sims K.B.;
RT "A protein with two immunoglobulin-like domains interacts with the
RT Norrie disease gene product norrin."
RL Am. J. Hum. Genet. 61:A158-A158(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Lockwood S.K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF464873; AAL69964.1; -.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000634; S/T_dehydrtse.
DR Pfam: PF00047; Ig; 5.
DR SMART; SM00409; IG; 5.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_1.
SQ SEQUENCE 1106 AA; 122019 MW; 80B52268DD90E7C6 CRC64;

Query Match 32.1%; Score 2213.5; DB 4; Length 1106;
Best Local Similarity 38.1%; Pred. No. 3.1e-139;
Matches 518; Conservative 152; Mismatches 305; Indels 383; Gaps 26;

QY 38 SNNPCIFGSPGAEGGGGDDLPDLAFLSQBELDESVMRLAINYDPLEKADETOAR 97
Db 5 SSHEFYDLSDMQEESEKNTDFPGLSAFLSQBELDESVMRLAINYDPLEKADETOAR 64
QY 98 KRL---SPDQMKHP-----NLSEFNFQDNPSP-----TSSKESPOEAK- 136
Db 65 SQIFSTSPALCEHPSHKETHKEGASRRP---QDNSTVPQPLAEKQTKSISPVSKRK 121
QY 137 -----RPOYC-----SETQSKV-----FLNKAADFIEELSS 163
Db 122 PAMSPLLTRPSYIRSLRAEKRAKTPSTNVKPTPHORKGGQSQQLCDRAANLIEELTS 181
QY 164 LFSKSHSKRIRPRACKNHNKSKLESQNVQMSNSSFSDLSERRERSVPPIPADTRDNE 223
Db 182 IFKAAPRNRRSPNG-----ESSPDSGYLSPKNQPSALLSASAS----- 220
QY 224 VNHALEQEAQKRAEQAASAAAGDTPPGSSP-----SSLYEPLGPPPRFTQ 273
Db 221 -QSPMEDQGEWEKRVKSPGARHCYQNDQDLAVPHNRKSHFPHSALHFP-----AAPRFIQ 275
QY 274 KLRSPVEGTRVQLDCIVVGIPIPPQVRYWYCEGKELNESPDIHI-VQAGNLHSLTIAEAF 332
Db 276 KLRQEAQSVRYLRCVTRGNTPRVRFCEGKELHNTPDQIHCEGDLHTLIIAEAF 335
QY 333 EDTGRYSFCASNLYGTDTSDSAEIIYIGVSSSDSEGD-----PNKEMNRIQKPN----- 383
Db 336 EDDTGRTCLATNPSGSDTTSAEVFIEGASSTDSDESLSAFKSRAGAMPOAKKTTSVSL 395
```

```
QY 384 ---VSSPPT---TSAVIPPAVPAQHLVAQPRVATIOQCQSPNTNYLOGLDG-KPIIAAPVF 437
Db 396 TIGSSSPKTCGTTAVI-----QPLSVVQOVHSPSYLCRDPDCTTAYFPFVF 443
QY 438 TKMLNLSASEGOLVVFECRVKAGSPKVEWYREGTLIEDSPFRILQKKPRSMARP 497
Db 444 TKELQNTAVAEQGVVLECRVGRGAPPLQVQWFRQGSIQDSPFRILQKKPRSTAEPEEI 503
QY 498 CTLVIAEVEAFEDSGCTCTASNKYGTVSSIAQLHVRGNEDSLNNGLSHLSANSSTTLAAIE 557
Db 504 CTLVIAETFPEDAGIFTCSARNIDYGSATSTAQLVVTTSAN--TENCSTYESGESNNDHFQH 561
QY 558 POPSPHSEPPSVEQPPKPLEGVLVNHNPRSSRILGRVHFNLPDDDDKSGSEASSAGV 617
Db 562 FPPPPPILETSSLELASKPKPSEIQOVN-NPELGLSRAALQWQFNAAERE-----TSGV 613
QY 618 VTTRQTRPDSXQERFNGQATKT---PEPSF---PVKEPPPVLAQPKLDSTQLOQLHNOVL 671
Db 614 HPSR-----GVNGLINGKANSNKSILPTPAVLLSPTKEPPPLAKPKL----- 655
QY 672 LEQHQLQNPPSPSPKPEFPFMTVLNSNAPPAVTTSSXQVKAPSSQTFSLARPKYFFPSTN 731
Db 656 ----- 655
QY 732 TTAATVAPSSSPVFTLSSTPQTIIQRTVSKESLLVSHPSVQTKSPGLSIOQLNEPLPPGPT 791
Db 656 ----- 655
QY 792 PTPPPFTFSIPSGNQFQPCRVPIPVSPTRSIQNPVAFLSVLPSLPAIPTNAXMLPRS 851
Db 656 ----- 655
QY 852 APSMPSQGLAKNNTKSPQVNDONIRETKNAVIRDLGKKKITFSDVRPN----- 899
Db 656 -----GFPKKASRTARIADEEIQCTKDAVTDQLERLKRFRKEDLLNNGQPRITYEERM 708
QY 900 -----QOEYKISSFEORLANEIEFRLERTPVDESDDDEIQHDE 936
Db 709 ARLLGADSATVFNIOQPEEBETANQEVKVSCEQRLISEIYRLERSPVDESDEVQYGD 768
QY 937 IPTGKCIAPIDKRLKHFRTVEGSPVTTCKIVGIPVYKVGDKGKOISKRNEHCKMR 996
Db 769 VPVENGMAPFEFMKLKHKIFEGMPVFTCRVAGNPKPKIYFWFKDGKQISPKSDHYTIOR 828
QY 997 EGDGTCSLHTESTSDDDGNYTTMAANPQGRISCSGHLVQSLPIRSLTSAGQSH---- 1052
Db 829 DLIDGTCSLHTASTLDDDDGNYTTMAANPQGRISCTGRMLVQAVNQGRSPRSPSGHPHVR 888
QY 1053 --RGRSRVQERDKPELQERFRPHFLQAPGDMAHGRCLRCLCKVSGLPPELTWLLNG 1110
Db 889 RPRSRSRSDGDENEPIQERFRPHFLQAPGDLTVOEGKLCMDCKVSGLPTPDLISWLDG 948
QY 1111 QVLPDASHKMLVRETGVHSLLDPLDQORDAGTYKCIATNKTNQNSFSLSVVAKEVKK 1170
Db 949 KPRPDSAHKMLVRENGVHSLIIEPTVTSRAGIYCTIATNRAGNSFSLSVLVAKEAHK 1008
QY 1171 APVILEKLQNGVPEGHVPRLECRVIGMPPPVFWKKDNETIPTCTREISMHOOTTYGAC 1230
Db 1009 PPVIEKLQNTGVADGYVPRLECRVGLVPPQIEFWKKENESLTHSTDRVSMHQNHGYIC 1068
QY 1231 LLIQAPKSDAGWVTTLSAKNEAGIVSCTARLDIYAQWH 1268
Db 1069 LLIQATKEDAGWTVTSKNEAGIVSCTARLDVYISRH 1106

RESULT 4
Q9Y2J6
ID Q9Y2J6 PRELIMINARY; PRT; 772 AA.
AC Q9Y2J6;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
```

```
DE KIAA0992 protein (Fragment).
GN KIAA0992.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023209; BAA76836.1; -.
DR HSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF00047; Ig; 3.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00408; IGC2; 3.
KW Immunoglobulin domain.
FT NON_TER 1
SQ SEQUENCE 772 AA; 83643 MW; EE5EE2BC79C7492B CRC64;

Query Match 24.2%; Score 1666; DB 4; Length 772;
Best Local Similarity 45.2%; Pred. No. 7.2e-103;
Matches 361; Conservative 113; Mismatches 193; Indels 132; Gaps 20;

QY 621 RQTPDSXQERFNGQAKTPSPVKEPPVLA-----KPKLDSTQLQOLH 667
DB 7 RGKPDPT-----CPAGPRSLQLP---LAPDAESGSGRRRPEPR-DPLKLOQL 53

QY 668 NQVLEQHQ-LQNPPP---SSPKFEFF-----XMTVLNSNAPPAVT 704
DB 54 NQIRLEQEGARQPPAPRSAPSPPPPPPPAFPELAACTPPASPEMSALASAPAM- 112

QY 705 TSXKQVAPSSQTSFLARPKYFFSTNTAAT---VAPSSSPVFTLSPTQTIQTVSKES 762
DB 113 -----QSSGSFNYPARPKQFTAAQNLGPGASGHGTASSP--SSSLPSPMSTPRQFG 162

QY 763 LLVSHPSVQ-----TKSPGGLSIQNEPLP---GPTPTPTPTFTSPSQNPQPCVCS 813
DB 163 RAVPPPAQPGAPPEAPWSSSPSPPPPPPPVSPFAAFVDPVDFPLP-----PPPPPL 217

QY 814 PIP-----VSPTSRI---QNPVAFSLSVLPSLPAIPPTNAMXLPSPAPSMPQSGLAKN 864
DB 218 PSPQASHCSSPATRFHGQTPAFLSALLPSQPPPAVNALGLPK---GVTPAGFPKKA 274

QY 865 TKSPQPVNDONIRTKNAVTRDLGKKITFTSDVRPN----- 899
DB 275 SRTARIASDEEIOGTDAVIQDLERKLFKEDLLNNGQPRLTYYERMARRLLGADSATVF 334

QY 900 -----QOEYKISSEFORLWNEIEFLERTPVDESDEIQHDEIPTCKIAPIFDK 949
DB 335 NIQPEETANQYKVSCEORLISEIYRLERSPVDESDEVQYGVDPVENGMAPFEM 394

QY 950 RLKHFRTYTESPVFTTKIVGIPVKYWFKDGKQKQISKRNECHKMRREGDGTCSLHTEST 1009
DB 395 KLHKYKIFEGMPVFTTCRVAGNPKPKIYWFKDGKQKQISPKSDHYTIQRDLGTCSLHTAS 454

QY 1010 TSDDDGNYTMAANPQGRISCSGHLMVQSLPIRSLRITSAGOSH-----RGRSRQERDK 1063
DB 455 TLDDDGNYTMAANPQGRISCTGRLMYQAVNQGRSPSPSGHPHVRPRSRSDSDEN 514

QY 1064 EPLQEREFRRHFLQAPGDMVAHEGRCLRLCKVSGLPPELTLWLLNGQVLPDASHKMLV 1123
DB 515 EPIQEREFRRHFLQAPGDLTVQEGKCLRMCKCKVSGLPPTDLSWLDGKGPVPDPSAHHMLV 574

QY 1124 RETGVHSLLDPLTQTDAGTYKCIATNKTGQNSFSLSVVAVKEVKKAPVILEKLQNCGV 1183
```

```
DB 575 RENGVHSLIIEPTVRDAGIYTCIATNRAGQNSFSLVWAAKEAHPVPVFIKLTQNGV 634
QY 1184 PEGHPVRLERCVICMPPPPVYWKXNDTIPCTRERISHQDQTTGYACLLIQPAKKSAGW 1243
DB 635 ADGYPVRLERCVLGVPPQPIFWKKENESLTHSTDRVSMHQDNHGYICLLIQGATKEDAG 694
QY 1244 YTLSAKNEAGIVSCTARLDIYAQHHH--QIPPPMSVPSGSRYSGLTSKGLDIFSASSM 1301
DB 695 YTVSAKNEAGIVSCTARLDVYVYQHQSQSTKPKKVRPSASRYAALSDDGLDIIKAATQP- 753
QY 1302 ESTMVYSCSSRSVSEDEL 1320
DB 754 EANPSHLTLTALVESEDL 772

RESULT 5
Q9ET54
ID Q9ET54 PRELIMINARY; PRT; 492 AA.
AC Q9ET54;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Actin-associated protein palladin (Fragment).
GN 2410003B16RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/NIH;
RX MEDLINE=20391984; PubMed=10931874;
RA Parast M.M., Oley C.A.;
RT "Characterization of palladin, a novel protein localized to stress
RT fibers and cell adhesions.";
RL J. Cell Biol. 150:643-656(2000).
DR EMBL; AF205078; AAG00078.1; -.
DR HSP; P56276; ITLK.
DR MGD; MGI:191583; 2410003B16RIK.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00409; IGC; 3.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; IG_like; 1.
KW Immunoglobulin domain.
FT NON_TER 1
SQ SEQUENCE 492 AA; 55072 MW; BC59E5B3E3BAEBD5 CRC64;

Query Match 21.2%; Score 1460; DB 11; Length 492;
Best Local Similarity 55.7%; Pred. No. 2.2e-89;
Matches 274; Conservative 78; Mismatches 96; Indels 44; Gaps 5;

QY 872 NDNIRETKNAVTRDLGKKITFTSDVRPN----- 899
DB 2 SDEEIQGTCTAATQDLERKLFKEDLLNNGQPRLTYYERMARRLLGADSANVFNIQPEE 61

QY 900 ---QOEYKISSEFORLWNEIEFLERTPVDESDEIQHDEIPTCKIAPIFDKRLKHFV 956
DB 62 TAANQEKVVSCEORLISEIYRLERSPVDESDEVQDPVVENATAPFEMKLLHYKI 121

QY 957 TEGSPVFTTKIVGIPVKYWFKDGKQKQISKRNECHKMRREGDGTCSLHTESTTSDGDN 1016
DB 122 FEGMPVFTTCRVAGNPKPKIYWFKDGKQKQISPKSDHYTIQRDLGTCSLHTASTLDDGDN 181

QY 1017 YTIMAANPQGRISCSGHLMVQSLPIRSLRITSAGOSHRSRVRQERD----KEPLQERF 1070
DB 182 YTIMAANPQGRVSTGRLMYQAVNQGRSPSPSGHPHARRPRSRSDSDENEPQIERF 241

QY 1071 FRPHFLQAPGDMVAHEGRCLRLCKVSGLPPELTLWLLNGQVLPDASHKMLVRETGVHS 1130
```



```

Db 320 TKEDAGTYTSAKNEAGIVSCTARLDVYQWQQSQSTKPKKVRPSASRYAALSDDOGLDI 379
QY 1295 PSFASMESTWVYSCSSRSVVESEDL 1320
Db 380 KAATQP-EANPSHLTLNLTALVESEDL 404

RESULT 8
Q9CWW1 ID Q9CWW1 PRELIMINARY; PRT; 379 AA.
AC Q9CWW1; 2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 2410003B16Rik protein.
GN 2410003B16Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRIONIC STEM CELLS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010350; BAB26871.1; .
DR HSP; P56276; lTLK
DR MGD; MGI:1919583; 2410003B16Rik.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR Pfam; PF00047; Ig.
DR SMART; SM00409; Ig.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; IG_like; 1.
DR Immunoglobulin domain.
SQ SEQUENCE 379 AA; 42157 MW; 42A96BE0AAC8F25 CRC64;

Query Match 18.5%; Score 1271.5; DB 11; Length 379;
Best Local Similarity 61.5%; Pred. No. 6.1e-77;
Matches 233; Conservative 61; Mismatches 76; Indels 9; Gaps 4;

QY 950 RLKHFRVTEGSPVFTTCKIVGIPVKYVWFKDGKQISKNECHKMRREGDGTCSLHTEST 1009
Db 2 KLHKYKIFEGMPVFTTCRVAGNPKPKIYWFKDGKQISPRSDHYTIQDLDTCSLHATTS 61

QY 1010 TSDDDGNTYITMAANPQGRISCSGHLMVQSLPIRSR--LTSAGSHRGRSRVQRD-----K 1063
Db 62 TLDDDGNTYITMAANPQGRVSCVTCGLMVQAVNQGRSPSPSGPHARRPRSRSDSDEN 121

QY 1064 EPLQERFRPHFLQAPGDVMAHEGRCLRDCKVSGSLPPELTWLLNQPVLPDASHKMLV 1123
Db 90 TSTAELLVTAGTAPPNFSQRLQSMRTARQSGQVRLDVRVTGTPVVKFYRGDVEIQSSPD 149

```

```

Db 122 EPIQERFRPHFLQAPGDLTVQEGKLCRMCKVSGSLPTDLSWQDQKPIRPDSAHKMLV 181
QY 1124 RETGVHSLIDPLTORAGTYKCIATNKTSQSFSLSESVVAKVKKAPVILEKQNGV 1183
Db 182 RENGVSLLIEPTVSRDAGIYTCIATNRAGONSFNLELVAAKHAHRAVPVMEKQNGV 241
QY 1184 PEGHPVRLECRVIGMPPPVYVKKKNETIPTCTRERISMHQDQTTGYACLLIQAPKKS DAGW 1243
Db 242 ADGYPVRLECRVSGVPPPIFWKKENESLHSTERTVSMHQDNHGYICLLIQGATKEDAGW 301
QY 1244 YTLSAKNEAGIVSCTARLDIYAQHH--QTPPPMSVPSRSRYSGLSLTKGLDIFSAFSSM 1301
Db 302 YTVSAKNEAGIVSCTARLDVYQWQQSQSTKPKKVRPSASRYAALSDDOGLDIKAATQP- 360
QY 1302 ESTWVYSCSSRSVVESEDL 1320
Db 361 EASPSHLTLNLSGLVESEDL 379

RESULT 9
Q98918 ID Q98918 PRELIMINARY; PRT; 4162 AA.
AC Q98918; P87479;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Connectin/titin (Fragment).
GN TITIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST MUSCLE;
RX MEDLINE=96254045; PubMed=8660363;
RA Yajima H., Ohtsuka H., Kawamura Y., Kume H., Murayama T., Abe H.,
RA Kimura S., Maruyama K.;
RT "A 11.5-kb 5'-terminal cDNA sequence of chicken breast muscle
RT connectin/titin reveals its 2 line binding region.";
RL Biochem. Biophys. Res. Commun. 223:160-164(1996).
RN [2]
RP SEQUENCE OF 1767-1871 FROM N.A.
RX MEDLINE=96365546; PubMed=8769723;
RA Turnacioglu K.T., Mittal B., Sanger J.M., Sanger J.W.;
RT "Partial characterization of zeugmatin indicates that it is part of
RT the 2-band region of titin.";
RL Cell Motil. Cytoskeleton 34:108-121(1996).
DR EMBL; D83390; BAA11908.1; .
DR EMBL; U64829; AAC60019.1; .
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003606; Ig_MHC.
DR Pfam; PF00047; Ig; 33.
DR SMART; SM00408; IGC2; 14.
DR SMART; SM00410; IG_like; 18.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR Immunoglobulin domain.
FT NON_TER 4162 4162
SQ SEQUENCE 4162 AA; 464971 MW; DB0905C907686649 CRC64;

Query Match 12.6%; Score 871.5; DB 13; Length 4162;
Best Local Similarity 23.5%; Pred. No. 8.9e-49;
Matches 327; Conservative 193; Mismatches 507; Indels 367; Gaps 43;

QY 209 SSVPIPIPADTRDNEVNH-----LEQQEAKRREAEQAASEAAGG----DTTPGSSPS 257
Db 30 SGFPVPEVSVWRDQVLSAATLPGVQISFSDGRAKLVIPSVTEANSGRYTIQATNGSSQA 89

QY 258 SLTYEEPL---GQPPRTQKLSREVPEGTQVLDCTVIGIPPPQVRYCEGKELENSPD 314
Db 90 TSTAELLVTAGTAPPNFSQRLQSMRTARQSGQVRLDVRVTGTPVVKFYRGDVEIQSSPD 149

```

QY 315 IHVQAGNLHSLTAAFEEDTGRYSCFASNIYCTDSTSAEIIYEGVSSSDSECDPNKE- 373
 Db 150 FQILQEGDLSLIAEAYPEDSGTYSVNTNNGVRATSTAEILLIQ-----EEAVPAKKT 205
 QY 374 -----EMNRIQK-----PNEVSSPPTTS 391
 Db 206 KTIIVTAQISQTRQARKEIETHFDARSLSVEMVIEGAAQAOLPHKAPRMPPTSK 265
 QY 392 AVIPPAVPAQHLVAOPRVATIQOCQSPNYLOGLDGKPIIAAPVFTKMLQNLASSEGQL 451
 Db 266 SPPIPVITAKAQAQSPSPVRHVRAPTSPVRSV-----SPAGRI 314
 QY 452 VFECRYKAGSP--KVE-----WYREG--TLIEDSPDFRI----- 483
 Db 315 STSPIRPVKSPSIRKAAVQVTPGAELVLPWRQEGYSATAEAQMKETRVSTSAEIRTEER 374
 QY 484 -----IQ-----KKPRSMABEIECTLVIA-----EV 505
 Db 375 WEGRYGLQEQVTTISGAAGEVAAGAKAEVRKEPEKTPVPTVIATDKAKEQERISTAREEI 434
 QY 506 FAEDSGCFTCTASKYG-----TVSSSTAOLHVRGN----- 535
 Db 435 SARHEQVHVSHQIEAGKRAEAVATVVAADQARVRSWPETEODEYVYVKKKLEYGYKE 494
 QY 536 ---EDLSNNGSLHSANSTTNLAAIEPOPSPPHSEPPSV-----EQPPKPLEGLV-VNH 585
 Db 495 HAYKDHEAQAEHVAHVATKEVTVVPPKHIPAAEKKEHVHSTEIKRETEAKIEKTHIEH 554
 QY 586 NEPRSSRIGLRVHFN-----LEDDKGSSEASAEAGVVVTRQ-----TRPDSX 628
 Db 555 PRPRTAS-----PHFTVSKIAVPKPDHTYEVSVIAGSAMATLEKELSATSAAQKITKVPK 609
 QY 629 QERENGQATKPE--PSEFVKEPP-----PVLAKPKLDSTQLQHLNQVLEQHLQ 678
 Db 610 POLKPHVEKIKPSAPPQFFTEAAETVYKAHYDVETKKEVDVS-----IKGEAVREDHLL 665
 QY 679 NPPPS-----PKPEFFXMTVL-----NSNAPPAVFTSX 707
 Db 666 RKESEAKVETETARVPVPAEIPVPTLVWGLKKNKTVTEGESVILECHISGHPQPTVWYR 725
 QY 708 KQVKAQSSQTFSL-----ARKYFFPSTNTAATVAPSSPVFTLSSTPQ 752
 Db 726 EDYKIESMDFOITFKAGLARLVIREAFEDSGRFTCTATNKAGSVSTSCHLHVXVSEET 785
 QY 753 TIQRTYSKESLLVSHPSVQTK-----SPGGLSIQNEPLPP-----GTEP 792
 Db 786 ETRETISEKVVTEKSVETKQVVMEDVSAAAEVSGEVPVPPFFIRKPVVVKLIEGGS-- 843
 QY 793 TPPPFTFIPSGNQFQPCV--SPIVPSPTS----- 822
 Db 844 ---IIFECQVGGNPKPHVLWKKGGVPLTTGYRYKVSVKRETGECKLEISMFTADDAGEY 899
 QY 823 ---IQNPVAFVLSVLPSPALPPTNAXMLPRSPAPMSQGLAKNKTSPQ-----PV-- 871
 Db 900 TIVIRNKGESA---TVSLEEADYEAAYIKSQOEMMYQQTVAAYQEPVAVAEAPPISY 956
 QY 872 -NDNIRETKNAVIR-DLGKKTIFSDVRPNQOEYKISSFEQRLMNEIEFLERTPVD--- 926
 Db 957 GDFDKYEKEQALIRKKNADVTWVRFTVEDEEFHISFEERLKELEIRLIKTLDL 1016
 QY 927 ESDDIEQIHDEIPTGKCIAPIDFKLRHFRVTEGSPVFTTCKIYGPVYVWFKDGKQIS 986
 Db 1017 EEDGEENMIDISEAIGAGFDLRKLYRTFEGTVTFHCKTGYPLPKTAWYDKGRRI- 1075
 QY 987 KRNECHKMRREGDGTCSLHISTESTDDGNVYTIMAANPQGRISCSGHLMVQSL----- 1039
 Db 1076 RHGERYHMEVLQDGSASRLPVVLPPEDEGIYTVFASNMGNAICSAKLYVEVPAPTATPG 1135
 QY 1040 -----PFRSLTSAGQSHRGRSRVQERDKEPLQERFPRPHFLOAPGD 1081
 Db 1136 YMPGPEVMRRYRISPRSPSRSPARSPSPAR-RUDETDEGL-ERLYKPVVFLKPTS 1193

QY 1082 MVAHEGRCLRDCKVSGLPPELTWLLNGQVLPDPASHKMLVRETGVHSLIDPLTQDA 1141
 Db 1194 VKCSQGTAREDLKVVGRPMPTWYFHNQGVVNDYTHKIVIKEDGTQSLIIVPAMPEDS 1253
 QY 1142 GTYKCIATNTKQNSFSLSLSVAKYKAPVILEKLQNGVPEGHVPVRLCVRIGMPPPP 1201
 Db 1254 GEWAVIAQNRAGKASVSVTLSEAKEDLVRPRFVERLRNVSVKESGRHLHMAVKATGNPNP 1313
 QY 1202 VFVWKDNETI-PCRERISMHODTGTGACLLIOPAKKSDAGWYTLTSAKNEAGIVSCTAR 1260
 Db 1314 DIVLKNDDIIVPHKYPRII-EGTKGAAALNTESTARQDAAMTATATAINKAG--ROTR 1370
 QY 1261 LDIYAQWHHQIPPP 1274
 Db 1371 CKVNVVEVHAPEP 1384
 RESULT 10
 Q9IAR9
 ID Q9IAR9 PRELIMINARY; PRT; 2164 AA.
 AC Q9IAR9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Structural muscle protein titin (Fragment).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 RX MEDLINE=20084964; PubMed=10618168;
 RA Ayoub J.C., Turnacoglu K.K., Mittal B., Sanger J.M., Sanger J.W.;
 RT "Targeting of cardiac titin fragments to the z-bands and dense bodies
 of living muscle and non-muscle cells";
 RL Cell Motil. Cytoskeleton 45:67-82(2000).
 DR EMBL: AF159173; AAF35436.1;
 DR InterPro: IPR003598; Ig_c2;
 DR InterPro: IPR003600; Ig_c2;
 DR InterPro: IPR003006; Ig_Like.
 DR Pfam: PF00047; Ig; 10.
 DR SMART: SM00408; IGC2; 4.
 DR SMART: SM00410; IG_Like; 6.
 KW Immunoglobulin domain.
 FT NON_TER 2164 2164
 SQ SEQUENCE 2164 AA; 242239 MW; 725852C3888875B CRC64;
 Query Match 12.0%; Score 826; DB 13; Length 2164;
 Best Local Similarity 22.0%; Pred. No. 4e-46;
 Matches 351; Conservative 195; Mismatches 454; Indels 598; Gaps 49;
 QY 209 SSVPIPIATDRDNEVNH-----LEQOEAKRREAQAASEAAG-----DTPGSSPS 257
 Db 30 SGFVPEVSWRQVLSAATLPGVQISFSDGRAKLVPSTVTEANSGRYTIQATNGSQ 89
 QY 258 SLYVEEPL--GQPRPTOKLRSREVEPTGVQVLDCLVVGIPPPQVRYWYCEGKLENSPD 314
 Db 90 TSTAELLVTAGTAPPNESQRLOQSMTRQSGQVRLDVRVTGIPTPVVRYFGVGIQSSPD 149
 QY 315 IHVQAGNLHSLTAAFEEDTGRYSCFASNIYCTDSTSAEIIYEGVSSSDSECDPNKE- 373
 Db 150 FQILQEGDLSLIAEAYPEDSGTYSVNTNNGVRATSTAEILLIQ-----EEAVPAKKT 205
 QY 374 -----EMNRIQKNEVSSP-----PTTS 391
 Db 206 KTIIVTAQISQTRQARKEIETHFDARSLSVEMVIEGAAQAOLPHKAPRMPPTSK 265
 QY 392 AVIPPAVPAQHLVAOPRVATIQOCQSPNYLOGLDGKPIIAAPVFTKMLQNLASSEGQL 451
 Db 266 SPPIPVITAKAQAQSPSPVRHVRAPTSPVRSV-----SPAGRI 314


```
Db 863 TTVEREFVHLHGREAKVTETARPAPVEIPV-----TPPTLVSLGNVTVIEGESV 914
QY 719 SLARPKYFFPSTNTT----- 733
Db 915 TLECHISGYSPVTVWREDYQIESSIDFOITQSGIARLMIREAFEDSGRTCSAVNE 974
QY 734 AATVAPSS-----SPVFTLSSTPQTQRTVSKESLIVSHPSVQTKSPGGLSTQNEPLPP 787
Db 975 AGTVSTSCYLAVQVSEFEKETTAETKTEKRFVESRDVWMTDI---SLTEE--QA 1028
QY 788 GPTEPTPPF-----TFSIDSGNQFQRCV---SPIPVSTSR----- 822
Db 1029 GPGEPAAPYFITKPVVQKLVEGGSWFGCQVGNPKPHVYWKXSGVPLTTTCYRYKVSYNK 1088
QY 823 -----IQNVAFLSVLPALPAPPTNAMXLMPSRAPSMSQ 859
Db 1089 QTCEKLVISWTFADDAGEYTIIVRNKHGETSA---SASLEEADYELLMKSQOEMLYQ- 1144
QY 860 LAKNNTKSPQVNDNDIRET-----KNAVIR-DLGKKITFSDVRPNQOEY 903
Db 1145 ---TQVTAFAVQEPKVGETAAGFVSEYEKEQALIRKKMAKDTVVVVRTVVEDQEF 1199
QY 904 KISFEORLMNEFERLERIP-----VDESDDEIQHDPIPTGCKTAPFDKRLKHFVTEG 959
Db 1200 HISSFEERLIKEIYRIKTTLEELLEDEGEKMAVDISESEAVESGFDLRIKNYRILEG 1259
QY 960 SPVTCTKIVGIPVKYVWEKDGKQISKRNEHCKMRREGDGTCSLHTESTSDDGNYTI 1019
Db 1260 MGVTFHCKMSGYPLPKTAWYKDGRI-KHGERYQMDFLQDGRASLRIPVVLPEDEGIYA 1318
QY 1020 MAANPOGRISCSGHLMVQ-----SL-----PIR----- 1042
Db 1319 FASNIKGAICSGKLYVEAPLGPATYIPTLEPVSRISLSRPSVSPIRMSPARMS 1378
QY 1043 SRLTSACQSHRGS---RVQERDKEPQERFRPHFTQAPQDMVAHGRCLRCDCKVSL 1099
Db 1379 ARMSPARMSPARSGRLEETDESQ-ERLYKPFVFLKPVSEKCLGQATARFDLKVVG 1437
QY 1100 PPELTWLLAGQVLPDASHKMLVRETVHSLLDPLTORDAGYKCIATNKTGNSFSL 1159
Db 1438 PMETFWFHGQQIVNDYTHKVIKEDGTQSLIIVPATPSDSEGTWVVAQNRAGRSISV 1497
QY 1160 ELSVAKEVKKAPVLEKLQNGCYGPEGHVRLCECRVICMPPPVFYMKKDNETI-PCTRER 1218
Db 1498 ILTVEAVEHQVKPFVEKLNKVNKEGSQLEMKVRATGNPNPDIVLWLNKSDIIVPHKYK 1557
QY 1219 ISMHQDTTGACLLIQAPKSDAGWYTLAKNEAGIVSCTARLDIYAQWHQIIPPPMSVR 1278
Db 1558 IRI-EGTKGEAALKIDSTVSDQSAWYTATATINKAGRDTTRCKNVNEVEFAEPEPERKLII 1616
QY 1279 PGS 1282
Db 1617 PRGT 1620

RESULT 13
O97791
ID O97791 PRELIMINARY; PRT; 2000 AA.
AC O97791;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Titin (Fragment).
GN TTIN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SOLEUS SKELETAL MUSCLE;
RX MEDLINE=99034591; PubMed=9817758;
RA Gregorio C.C., Trombitas T., Kolmerer B., Stier G., Granzier H.,
```

```
RA Kunke K., Suzuki K., Obermayr F., Herrmann B., Sorimachi H.,
RA Labelit S.;
RT "The N terminal of titin spans the Z-Disc. Its interaction with a
RT novel 19 kDa Ligand (T-cap) is required for sarcomeric integrity.";
RL J. Cell Biol. 143:1013-1027(1998).
DR EMBL; Y18102; CAA77028.1; -.
DR HSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 9.
DR SMART; SM00408; IGC2; 4.
DR SMART; SM00410; IG_Like; 4.
KW Immunoglobulin domain.
FT NON_TER 2000 2000.
SQ SEQUENCE 2000 AA; 222702 MW; D157B54C33D5B68 CRC64;

Query Match 11.5%; Score 793; DB 6; Length 2000;
Best Local Similarity 21.0%; Pred. No. 5.7e-44;
Matches 339; Conservative 193; Mismatches 497; Indels 584; Gaps 46;

QY 198 SFDLSERRERSSVPIPIPADTRDRNEVNHAEQQEAKRREAEQAASEAAGDITPGSSPS 257
Db 57 SFS-----GRAKLTIPIAVTKANSGRYSL-----RATNGSGQATSAE 94
QY 258 SLVYEEPLGPPREFTKQLRSREVPEGTRVQLDCIVVGIPPPQVRYWYCEGKELENSPDIHI 317
Db 95 LLVTAET--APPNFVQLQSWTVRQGSQVRLQVRVTGPTPVVAFYKRDGAEIQSLDFQI 152
QY 318 VQAGNLHSLTAAFEEDTGRYSFASNIYQTDSTSAEIIYEG-----VSSDSEGD 369
Db 153 SQEGDLYSLLIAEAYPEDSGTYSVNAATNSVGRATSTAEALLVQGEVVPAAKTKTIVSTAQ 212
QY 370 PNKEEMRIQK-----PNEVSSPPTTSIVIPPAVP 399
Db 213 ISETROTRIEKKIDBAHFDARSIAIVEMVVDGATQQLPHKTPPRIPPKRSRSTPPSIA 272
QY 400 QAQHLVAQPRVATIQCCQSPNTNYLQGLDGPI-----IAAPVFTKMLQ 443
Db 273 AKAQARQSPSPRHSPSPVRHVRAATPPSPRSVSPAGRISTSPIRSKSPLLVRKTQT 332
QY 444 LSASEGOLVFECEKRVKGAESPKEVYREG-----TLIEDSPDFRI----- 483
Db 333 TTMATPEV-----PPD---WKQEGYVASSTEAEEMRETTWTSSTEIRREERWEGRY 380
QY 484 -LQKKPRMAEPPEICILVIAEVAEDSGCCTCTA-----SNKYGVSSI--AQLHVRG 534
Db 381 GLOEQVITGAAGAAAVSTSAAFAGLGSWFATGAKEVKQADKSAAVATVVAADVMAV 440
QY 535 NEDLSNGLSHSANTTNLAAIEPOPSPPHSEPPSV-----EQPK----- 575
Db 441 REPVISAVEQTAQRTTTTAVHIQPAEQVRKEAKIAVKKVVAADKAKEQELKSRPREI 500
QY 576 --PKLEGVLVNHNPR-----PVKEP-----PPVLAKPK---LDSTQL 663
Db 501 ITTKQEQMHITHEQLRKETEKAFFVKVVISAAKAKEQETRITGETITTKQEQKITQETMM 560
QY 590 -----SSSRIGL---RVHFNLPEDDKGSEASSEA-----GVV 618
Db 561 KETRKTVPVKIVATPKVKEQDLVSRREGISTKREQVQITQERMKRKEAETALSTIATA 620
QY 619 TTRQTRPDSXQERFNGQATK-----T 639
Db 621 TAKATEQETVLTREGMATROEQIQVTHGKVGCGKAEAVATVVAADVQARVREPREG 680
QY 640 PEPSP-----PVKEP-----PPVLAKPK---LDSTQL 663
Db 681 PEDSYAQOTTLEYGYKEHISATKVAEHPKRPASEPHVVPKAVKPVVQIAPSETHIKTTEQ 740
QY 664 QQLH-----NOVLLEQ---HQLQNPPSPSPKPEFPFXMTVLN-----SNAPPV 703
Db 741 VGMHISQIKKTTDVTSLRVLHVDKRRPTASPH---FTVSKISVPKTEHGEYASTAGSAI 797
```


Db	633	ITTKREOVQITQBCKMRKEAKETALSTAFIATATAKAEQETILTRTETMATRQEQIQVTHGK	632
Qy	517	-----ASNKYGTVSSIAQLHVRGNEDLSNNGSLHSANST-----TNLAAIEPQSP	562
Db	693	VDVGKKAEEAVATAVAQDQARVREPREPCHLEESYAQOTTLEYGYKERISAAKVAEPPQR	752
Qy	563	PHSEPPSVVEQPPKKL-----EGVLVNINEPRSS	591
Db	753	PASEPHVYVRAVKPRVIOAPSETHIKTTDQKGMHISSQIKKTTDLTTTERLHVHVDKRPRTA	812
Qy	592	SRIGLRVHF-----NLPEDDKGEASSEAGVYVTRQ-----TRPD	626
Db	813	S-----PHFTVSKISVPKTEHGEYASIAGSATATLOKELSATSQAQITKSVKAPTWKPS	867
Qy	627	SXQERFNGQATKTPESFPVKPEPPVYLAKP-----KLDSTOLOQLHNOVL-----LE	673
Db	868	ETRVR-----AEPPLPQFPADPTDPTYKSEAGVEVKVGVISGTIVREERFEVLHGRE	923
Qy	674	QHQLONQPPSPKPEFFPMATVLNSNAPPAVTTTSXQKVAPSQSFSLARPKYFPFPTNTT	733
Db	924	AKVTETARVAPVIEIV-----TPPLTVSGLKNVTVIEGESVTTLECHISGVPSPTVT	975
Qy	734	-----TFSPISGNQFQPCV---SPIDVSTPSR-----AATVAPSS-----S	742
Db	976	WYREDYQIESIDFQITFQSGTARLMIREAFADSGRFTCSAVNEAGTVTSYCLAVOVVS	1035
Qy	743	PVFTLSSTPOTIQTWYSKESLIVSHPSVQTKSPGSLSIQNEPLPPGTEPTTPPPF	797
Db	1036	EEPEKETTAVTERFTTEERKEFVESRDVVMTDI-----SLTEE--QAGPGCPAAPYFITKPV	1089
Qy	798	-----TFSPISGNQFQPCV---SPIDVSTPSR-----822	
Db	1090	VQKLVEGSGVYFCQVGGNPKPHVYWKKSGVPLTTGYIKYKYSYNKQGTCECKLVISMTFAD	1149
Qy	823	-----IONPVAELSSVLPSPALPPNNAKMLPSRAPSMPGSLAKKNTKSPQPVND	874
Db	1150	DAGEYITVWANKGETISA--SASLLEEADYELLKMSOEMLYQ-----TQVTAFFQEP	1200
Qy	875	NIRET-----KNAVIR-DLGKKITFSDYRPNQOEYKISSFEQRLMNIIEF	918
Db	1201	KVGETAPGFVYSEYEKEQALIRKKMAKDTVVVVVTVYVEDQBFISSPEERLIKEIY	1260
Qy	919	RLRTP-----VDSDEIQHDEIPTGCKTAPTFDKRLKHFRVTGSPVTFCKIVGTPVP	974
Db	1261	RIIKTLELLEEDGEEKMAVDISEASEYVESGFDLRIKRYILEGMVGYTFCKMSGYPLP	1320
Qy	975	KVYFKDGKQIKSRNRHCKMRREGDTCLSHTESTSDDDGNYITMAANPOGRISCSGHL	1034
Db	1321	KIAWYKDGKRI-KHGERYQMDFLQDGRASLRTPVVLPEDEGIYATAFNKGNACISGKL	1379
Qy	1035	MVQ-----SL-----PIR-----SRUTSAGQSHGRS-	1056
Db	1380	YVEPAAPLGAPTYIPTLEPVSRISLSRPSRSKPSIRMSPARMSPARMSPARMSPARMS	1439
Qy	1057	--RVQERDKPELQERFRHFHQAPGDWVAHGRCLRCDCKVSGLPPPELTWLLNGOPVL	1114
Db	1440	GRLEETDESQLE-ERLYKPVFLKPVSEKFCLEGTAREDLKVGVRPMPETFWFHDGQOIV	1498
Qy	1115	PDASHKMLVRETGVHSLLDPLTORDAGNYKCIATNKTGNSFSLSLSWAKEVKKAPVI	1174
Db	1499	NDYTHKVVIEEDGTQSLIIVPATPSDGEWTVVAONRAGRSISVILLTVEAVEHQVKPMF	1558
Qy	1175	LEKIQNCVGPGEHPVRLECRVIGMPPPVYWKKNDETI-PCTRERISMHQDQTTGYACLLI	1233
Db	1559	VEKLNKVNIREGSOLEMKVYRATGNPNPDIVWLKNSDIIIVPHKYPKIRI-EGTKGEALKI	1617
Qy	1234	QPAKKSAGWYITLSAKNEAGIVSCSTARLDIYAQWHHQIHPDPMWSVRPSGS	1282
Db	1618	DSTVQSDSANYTATNKAIGDRTTRCKGNVVEVEFAEPPEPPEPKLIIPRCT	1666

RESULT 15

Q8WZ42	PRELIMINARY;	PRT; 34350 AA.
ID	Q8WZ42	
AC	Q8WZ42;	
DT	01-MAR-2002 (TrEMBLrel. 20, Created)	
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Titin.	
GN	TTN.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20309627; PubMed=10850961;	
RA	Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereousse F.,	
RA	Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,	
RA	Granzier H., Labeit S.;	
RT	"Series of exon-skipping events in the elastic spring region of titin	
RT	as the structural basis for myofibrillar elastic diversity.";	
RL	Circ. Res. 86:1114-1121(2000).	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21573839; PubMed=11717165;	
RA	Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,	
RA	McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,	
RA	Labeit S.;	
RT	"The complete gene sequence of titin, expression of an unusual -700	
RT	kDa titin isoform and its interaction with obscurin identify a novel	
RT	Z-line to I-band linking system.";	
RL	Circ. Res. 89:1065-1072(2001).	
RL	EMBL; AJ277892; CAD12456.1; -.	
DR	InterPro; IPR000282; CytoK_receptor_2.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR000577; FGGY_kin.	
DR	InterPro; IPR003961; FN_III.	
DR	InterPro; IPR001092; HLH_basic.	
DR	InterPro; IPR003599; Ig.	
DR	InterPro; IPR003598; Ig_c2.	
DR	InterPro; IPR003006; Ig_MHC.	
DR	InterPro; IPR003596; Ig_v.	
DR	InterPro; IPR002016; Peroxidase.	
DR	InterPro; IPR004168; PPAK_motif.	
DR	InterPro; IPR002290; Ser_thr_pkinase.	
DR	Pfam; PF00041; fn3; 132.	
DR	Pfam; PF00047; Ig; 146.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF02818; PPAK; 53.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00060; FN3; 133.	
DR	SMART; SM00409; IG; 167.	
DR	SMART; SM00408; IGC2; 148.	
DR	SMART; SM00406; IGV; 23.	
DR	SMART; SM00220; S_TKC; 1.	
DR	SMART; SM00219; TyrcK; 1.	
DR	PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_1.	
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.	
DR	PROSITE; PS00290; IG_MHC; UNKNOWN_1.	
DR	PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.	
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.	
SQ	SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7C5E58A CRC64;	

```

Query Match      11.3%; Score 780.5; DB 4; Length 34350;
Best Local Similarity 20.0%; Pred. No. 2e+41;
Matches 333; Conservative 205; Mismatches 488; Indels 643; Gaps 42;

QY 198 SFSLSRERRSSVPIPADTRDNEVNHLEQQEAKRREAEQAASEAGGDTPTPGSSPS 257
      ||||| : ||| || : ||| : ||| : ||| : ||| : ||| : ||| :
Db 57 SFSF-----GRAKLTPIAVTKANGSRYSL-----KATNGSGQATSAE 94

QY 258 SLYEELPGQPPRFQKILRSREVPGSTRVQLDCLVVGIPPPQVRWYCSGKLENSPDTHI 317

```

Db 95 LLVKAET--APPNFVQRLQSMVTRQSGOVRLQVRVTGIPTPVVKFYRDGAIEIQSLDFQI 152
Qy 318 VQAGNLHSLTAIAEAFEDTGRYSCFASNIYGTDSIAEIIYEG-----VSSSDSEGD 369
Db 153 SOEGDLISLLIAEAYPEDSGTYSYNATNSVGRATSAELLVOGEEVPAKTKIIVSTAQ 212
Qy 370 PNKEEMRIQK-----PNEVSSPPTTSVAIPPAVP 399
Db 213 ISESRQTRIEKKIEAHFDARSATVEMVIDGAAGQQLPKHTPHRIPPKPKSRSPTPPSIA 272
Qy 400 QAQLHVAQPRVATTOQCOSPNTNYLOGLDGKPI-----IAAPVETKMLQN 443
Db 273 AKAQARQQSPSPRHSFSPVHRVRAPTSPSVRSVSPAARISTSPIRSVPSPLLMRRTQA 332
Qy 444 -----LSASEGOL-----VVFECEVRKVGAP 462
Db 333 STVATGPEVPPPKQEGVAVSSSEAEEMRETTLTSTQIRTEEREGRYGVOEQVTISGAA 392
Qy 463 S----- 463
Db 393 GAAASVSASAYAAEAVATGAKEVKQADKSAAVATVVAADMARVREPVISAVEQTAQR 452
Qy 464 ----- 463
Db 453 TTTTAVHIQPAQOVRKEAEKTAFTKVVAADKAKEQELKSRTKEVITTKQEQMHVTHEQ 512
Qy 464 -----PKV-----EWYRE----- 471
Db 513 IRKETETFPVKVISAAKAKEQETRISEITKKOKQVTOEAIHQETETIAASWVVVATA 572
Qy 472 -GTLIEDSP----- 479
Db 573 KSTKLETVPGAQEBETTQOQOMHLSYEKIMKETRTVVVPKIVATPKVKEODLVSRGREG 632
Qy 480 -----DFRILQKPRMAPEEETCTLVIAEVAEDSGCFTCT----- 516
Db 633 ITTKREQVQITQEKMRKEAEKALTSTIAVATAKAKEQETILRTRETMATROEQIQVTHGK 692
Qy 517 -----ASNKYGTVYSSIAQLHVRGNEDSLNSGSLHSANST-----TNLAAETPOPS 562
Db 593 VDCKKKAENATVVAADQARVREPDPGPHLEESYAQOTTLEYGYKERISAAKVAEPPQR 752
Qy 563 PHSEPPSVEQPPKPL-----EGVLVNHNEPRSS 591
Db 753 PASEPHVVPRAVKPRVIAQAPSETHIKTTDQGMHISQIKKTTDLTTERLVHVDKRPRTA 812
Qy 592 SRIGLRVHF-----NLPEDDKGSEASSEAGVVTTRO-----TRPD 626
Db 813 S-----PHFTVSKISVPKTEHGYEASIAIATLOKELSATSQAQITKRSVKAPTVPKS 867
Qy 627 SXQERFNGOATKTPERSFPVKPPVLAKEP-----KLDSTOLQQLHNOVL--LE 673
Db 868 ETRVR-----AEPTLPQOPFPADTPTDTKSEAGVEVKKEVGSITGTIVTERFEREVLHGHE 923
Qy 674 QHQLQNPPSPKPEFPFXMTVLNSNAPPVAVTTSXQVKAPSSQTFSLARPKYFPFSTNTT 733
Db 924 AKVTETARVPAPVEIPV-----TPPTLVSGLKNVTVIBGESVTLLECHISGYPSPTVT 975
Qy 734 -----AATVAPSS-----S 742
Db 976 WYREDYQIESIDQITFQSGIARLMIREAFEDSGRFTCSAVNEAGTVSTSCYLAVQVS 1035
Qy 743 PVFTLSSTPOTIORTVSKESLLVSHSPSVQKSGGLSTQNEPLPPGTEPTPPPF----- 797
Db 1036 EEFKEKETAVTEKETTEKREKRESDVNDT-----SLTEB--QAGGEPAAPIYITKPV 1089
Qy 798 -----TFSIPSGNQFQRCV-----SPIVSPSTR----- 822
Db 1090 VOKLVEGSGVFGCGVGNPKPHVYKKGSGVPLTGTGYRYKVSYNKQGTGECKLVISMTFAD 1149
Qy 823 -----IQNPVAFLSVLPSPAIPTTNAMXLPRAAPSQGLAKKNTKSPQPVND 874

Db 1150 DAGETTVIVRNKHGETSA---SASLLEADYELLMKSQOEMLYQ-----TQVTAFTVOEP 1200
Qy 875 NIRET-----KNAVIR-DLGKKITFSDVRPNQOQYKISSFEQRLMNIIEF 918
Db 1201 KVGETAGFVYSEYEKEQEALIRKKMAKDTVVVTYVEDQEFHISSEERLIKIEY 1260
Qy 919 RLERTP-----VDESDDIEQHDEIPTGKCIAPIFDKRLKHFRVTEGSPVTTCKTIVGIPVP 974
Db 1261 RIITKTTLEELLEEDGEBKMAVDISESEAVESGFDLRKKNYRILEGMGVTFHCKMSGYPLP 1320
Qy 975 KYVWFKDGKQISKNEHCKMRREGDGTCSLHIESTTSDDDGNVTYMAANQGRISCSGHL 1034
Db 1321 KIANYKDGKRI-KHGERYQMDFLQDGRASLRIPVVLPEDEGIYTAFAFNKGNACSGKL 1379
Qy 1035 MVQ-----SL-----PIR-----SRLTSAGQSHRGRS- 1056
Db 1380 YVEPAAPLGAPTYIPTLEPVSRIRSLSPRSVSRSPIRMSPARMSPARMSPARMS 1439
Qy 1057 --RVQERDEKPELOERFPRPHFLQAPGDMVAHEGRCLRDCKVSGLPPPELTWLLNGQPV 1114
Db 1440 GRRLEETDESQ-ERLYKPVFVLKPVSKLEGOTARFDLKVVGGRMPETFWPHDGOQIV 1498
Qy 1115 PDASHKMLVBRETGVHSLLDPLTORDAGTYKCIATNKTGQNSFSLSVVAKEVKKAPVI 1174
Db 1499 NDYTHKVVIKEDGTQSLIIVPATPSDGEWTVVAQNRAGRSSISVILTVEAVEHQVKPMF 1558
Qy 1175 LEKLONGCVPEGHVPVRLCECRVIGMPVPVYVWKDNETI-PCTREIRISMHODTTGYACLLI 1233
Db 1559 VEKLNKYNIKEGSOLEMKVRATGNPNPDIVWLKNSDIIVPHKYPKIRI-EGTKGEAALKI 1617
Qy 1234 QPAKKSADAGWYTLISAKNEAGIVSCTARLDIYAQWHHQIIPPMPSVVRPSGS 1282
Db 1618 DSTVSQDSAWYTATAIKAGRDTTTRCKNVNVEVEFAEPEPERKLIIPRGT 1666

Search completed: November 30, 2002, 19:06:18

Job time : 53 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 18:50:40 ; Search time 37 Seconds
(without alignments)
4753.806 Million cell updates/sec

Title: US-09-818-990b-2
Perfect score: 6890
Sequence: 1 MQDDIEASTISQLRESY.....MESTWVSCSSRVESDEL 1320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				A_Geneseq_101002:*
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*			
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*			
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*			
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*			
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*			
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*			
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*			
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*			
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*			
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*			
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*			
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*			
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*			
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*			
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*			
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*			
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*			
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*			
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*			
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*			
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*			
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*			
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	2560	37.2	507 22 AAB92617 Human protein sequ
2	2174	31.6	408 22 AAG10611 Novel human diagno
3	1477	21.4	544 21 AAB53464 Human colon cancer
4	1421.5	20.6	433 21 AAB42762 Human ORE2526
5	1394	20.2	264 22 ABB10330 Human cDNA SEQ ID
6	1380	20.0	261 22 AAU17989 Human immunoglobul
7	797.5	11.6	26926 22 AAU05396 Human titin (conne
8	729	10.6	498 21 AAY69234 Amino acid sequenc
9	724.5	10.5	159 22 ABB10509 Human cDNA SEQ ID
10	724.5	10.5	159 22 AAU18053 Human immunoglobul

11	656	9.5	130	21	AAB40875	Human ORFX ORF639
12	556	8.1	1953	23	AAU84351	Protein MYLK diffe
13	545.5	7.9	4796	22	ABB58665	Drosophila melanog
14	537	7.8	5635	23	ABF60991	Novel human protei
15	531.5	7.7	4412	21	AAV53666	Sequence gi/101742
16	475	6.9	3117	21	AAV53667	Sequence gi/332818
17	472	6.9	6815	22	ABG66811	Drosophila melanog
18	462.5	6.7	1240	22	ABG03933	Novel human diagno
19	446.5	6.5	201	23	AAU74341	Human cytoskeleton
20	426	6.2	7107	22	ABB58144	Drosophila melanog
21	416.5	6.0	2597	21	AAV53664	Mechanical stress
22	416.5	6.0	2597	23	AAAB47933	Rat OCP. Rattus r
23	410	6.0	2587	23	AAAB47933	Human OCP. Homo s
24	408.5	5.9	2387	21	AAV53665	Mechanical stress
25	392	5.7	4393	22	ABAB31889	Amino acid sequenc
26	382	5.7	4436	22	ABG23265	Novel human diagno
27	390	5.7	1853	21	AAV53668	Protein 608 sequen
28	388	5.6	1114	22	ABG66628	Drosophila melanog
29	387	5.6	2993	22	ABG23884	Novel human diagno
30	384	5.6	5266	22	ABG08561	Novel human diagno
31	383	5.6	3263	22	ABB67210	Drosophila melanog
32	381	5.5	2286	22	AAAB65635	Novel protein kina
33	381	5.5	2380	23	AAE19160	Human kinase poly
34	377	5.3	913	22	ABE59114	Drosophila melanog
35	367.5	5.3	162	23	ABP07380	Human ORFX protein
36	353.5	5.1	2542	22	ABE71137	Drosophila melanog
37	353	5.1	2016	22	ABB66424	Drosophila melanog
38	339	4.9	1336	23	AAAB47771	Human thrombospond
39	334.5	4.9	2596	22	ABAB30569	A splice variant o
40	332	4.8	681	22	ABB57982	Drosophila melanog
41	321.5	4.7	1297	20	AAV13565	C. elegans ROBO po
42	321.5	4.7	1297	20	AAV08403	C. elegans ROBO pr
43	318.5	4.6	1395	20	AAV13563	Drosophila Robo 1
44	318.5	4.6	1395	20	AAV08401	Drosophila sp. ROB
45	317	4.6	1465	23	AAE16625	Human M-protein en

ALIGNMENTS

RESULT 1	
AAAB92617	
ID	AAAB92617 standard; Protein; 507 AA.
XX	
AC	AAAB92617;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human protein sequence SEQ ID NO:10896.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 10896; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 507 AA;

Query Match 37.2%; Score 2560; DB 22; Length 507;
Best Local Similarity 100.0%; Pred. No. 3.8e-145;
Matches 494; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQDDIEASTSISOLLRESYLAETRRHGNERSRAEPSSNPCHFSGSPGAAEGGGQDDL 60
Db 1 MQDDIEASTSISOLLRESYLAETRRHGNERSRAEPSSNPCHFSGSPGAAEGGGQDDL 60
QY 61 PDLAFLSQEELDESVALRLAINDPLEKADETQAKRLSPDQMKHSPNLSFEPNFCQD 120
Db 61 PDLAFLSQEELDESVALRLAINDPLEKADETQAKRLSPDQMKHSPNLSFEPNFCQD 120
QY 121 NPSPTSSKESQAKRPQYCSQTSQKVKFLINKAADFIEELSLFKSHSKRIRPRACKN 180
Db 121 NPSPTSSKESQAKRPQYCSQTSQKVKFLINKAADFIEELSLFKSHSKRIRPRACKN 180
QY 181 HSKLESQNKVMQENSSFDLSRRERSVPIPIADTRDNEVNHAEQOAKRRAEQ 240
Db 181 HSKLESQNKVMQENSSFDLSRRERSVPIPIADTRDNEVNHAEQOAKRRAEQ 240
QY 241 AASEAAGDTPGSSPSLYEEPLGQPPRPTOKLRSREVPETGRVOLDCTVIGIPPQV 300
Db 241 AASEAAGDTPGSSPSLYEEPLGQPPRPTOKLRSREVPETGRVOLDCTVIGIPPQV 300
QY 301 RWYCEGLENSPDIHIVAGNLHSLTIAFAFEDTGRYSCFASNIYGTDSAEIYIEG 360
Db 301 RWYCEGLENSPDIHIVAGNLHSLTIAFAFEDTGRYSCFASNIYGTDSAEIYIEG 360
QY 361 VSSSDSGDPNKEEMNRIOKPNVSSPTTSVATPPAVPOAHLVAOPRVATIQCCQSPT 420
Db 361 VSSSDSGDPNKEEMNRIOKPNVSSPTTSVATPPAVPOAHLVAOPRVATIQCCQSPT 420
QY 421 NYLQGLDGKPIIAAPVFTKMLQNLASSEGOLVFECKVGAPSPKVEWYREGTLIEDSPD 480
Db 421 NYLQGLDGKPIIAAPVFTKMLQNLASSEGOLVFECKVGAPSPKVEWYREGTLIEDSPD 480
QY 481 FRILQKKPRMAEP 494
Db 481 FRILQKKPRMAEP 494

RESULT 2
ID ABG10611 standard; Protein; 408 AA.
XX
AC ABG10611;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #10602.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS74798.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 40970; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 408 AA;

Query Match 31.6%; Score 2174; DB 22; Length 408;
Best Local Similarity 99.5%; Pred. No. 3.7e-122;
Matches 406; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 913 MNEIEFLRTPVDESDEIQHDEIPTGKCTAPIFDKRLKHFRTGSPVFTCKIVGIP 972
Db 1 MNEIEFLRTPVDESDEIQHDEIPTGKCTAPIFDKRLKHFRTGSPVFTCKIVGIP 60
QY 973 VPKYWFKDGKQISKRNECHKMRREGDGTCSLHSTESTSDDGNYTMAANPQGRISCSG 1032
Db 61 VPKYWFKDGKQISKRNECHKMRREGDGTCSLHSTESTSDDGNYTMAANPQGRISCSG 120

thrombosis; contraceptive.
XX Homo sapiens.
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC76971.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 4239-4240; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX Sequence 433 AA;
Query Match 20.68; Score 1421.5; DB 21; Length 433;
Best Local Similarity 61.08; Pred. No. 4.7e-77;
Matches 261; Conservative 69; Mismatches 89; Indels 9; Gaps 3;
QY 901 QEYKISSFEORLMNEIFRLERTPVDSDDIOHDEITGKCIAPIDKRLKHFRTVEGS 960
DB 7 QEYKVSCEQRLSEIEYRLERSPVDSGGDEFTYGDVPVNGMAFFEMKLKHKRIFEGM 66
QY 961 PVTFTCKIVGIPVKYWFKDQKISKRNECHKMRREGDGTCSLHIESSTTSDDDGNYTIM 1020
DB 67 PVTFTCRVAGNPKKIYWFKDQKISPKSDHYTIQRDLDTGTCSLHTTASTLDDGNYTIM 126
QY 1021 AANPQGRISCSGLHWQSLPRTSRSLTSAQOSH-----RGRSRVQERDKPELQERFRPH 1074
DB 127 AANPQGRISCTGRMLVQAVNGRSPSPSGPHVRRPRSRSDSGDENEPQERFRPH 186
QY 1075 FLOAPGDMWAEGLRCLDCKVSLPPPELTWLLNGOPVLDPDASHKMLVRETGVHSLIID 1134
DB 187 FLOAPGDLTVOEGKLCRMDCVKSLPTPDLNSQLDGKVPVSDSAHKMLVRENGVHSLIIE 246
QY 1135 PLTORDAGTYKCIATNKNTGONSFSLELSVVAKEVKAPVILEKLQNCVGPBGHPVLECR 1194

Db 247 PVTSRDAGIVTCIATNRAGQNSFSLVWAAKEAHKPPVFTEKLTQNTGVADGYPVLECR 306
QY 1195 VIGMPPPVFWKKDNETIPCTRRERISMHODTTGYACALLIQPAKSDAGWYTLSSAKNEAGI 1254
Db 307 VLGVPFPQIFWKKENESLTHSTDRVSMHODNHGYICLLIQGATKEDAGWYTVSAKNEAGI 366
QY 1255 VSCARLDIYAOWHH--QIPPPMSVRPSSGSRYSGLTSKGLDIFSAFSSMESTMYVSCSSR 1312
Db 367 VSCARLDVYTOHQHQQSQSTKPKKVRFSASRYAALSQGLDIKAFQP-EANPSHLTLNT 425
QY 1313 SVESDEL 1320
Db 426 ALVESEDL 433
RESULT 5
ABB10330
ID ABB10330 standard; Protein; 264 AA.
XX AC ABB10330;
XX 10-JAN-2002 (first entry)
XX Human cDNA SEQ ID NO: 638.
XX Human; gene therapy; neural disorder; immune system disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation.
XX Homo sapiens.
XX WO200154474-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01349.
XX 31-JAN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180628P.
XX 02-MAR-2000; 2000US-184664P.
XX 16-MAR-2000; 2000US-186350P.
XX 17-MAR-2000; 2000US-189874P.
XX 18-APR-2000; 2000US-190076P.
XX 19-MAY-2000; 2000US-198123P.
XX 07-JUN-2000; 2000US-205515P.
XX 28-JUN-2000; 2000US-209467P.
XX 30-JUN-2000; 2000US-215135P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216880P.
XX 11-JUL-2000; 2000US-217487P.
XX 14-JUL-2000; 2000US-217496P.
XX 26-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225213P.
XX 14-AUG-2000; 2000US-225214P.
XX 14-AUG-2000; 2000US-225266P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225268P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225447P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 18-AUG-2000; 2000US-225759P.
XX 18-AUG-2000; 2000US-226279P.
XX 22-AUG-2000; 2000US-226681P.
XX 22-AUG-2000; 2000US-226686P.
XX 22-AUG-2000; 2000US-227182P.

PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 03-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234937P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.

PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
DR N-PSDB; ABA06552.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
PS Claim 11; SEQ ID NO: 638; 859pp + Sequence Listing; English.
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
SQ Sequence 264 AA;
Query Match 20.2%; Score 1394; DB 22; Length 264;
Best Local Similarity 99.6%; Pred. No. 1.1e-75;
Matches 261; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1059 QERDKPELQERFRPHFLQAPGDMVAHEGRLCRLDCKVSLPPPELTWLLNGQVLPDAS 1118
Db :|||||
3 RERDKPELQERFRPHFLQAPGDMVAHEGRLCRLDCKVSLPPPELTWLLNGQVLPDAS 62
QY 1119 HKMLVRETGVHSLIDPLTORDAGTYKCIATNKTGQNSFSLSVAKVKAIVILEKL 1178
Db :|||||
63 HKMLVRETGVHSLIDPLTORDAGTYKCIATNKTGQNSFSLSVAKVKAIVILEKL 122
QY 1179 QNCGVPEGHVRLRCRVIGMPVPVYWKKNDETICTPCTRERISMHQDTTGACLLIQPAKK 1238
Db :|||||
123 QNCGVPEGHVRLRCRVIGMPVPVYWKKNDETICTPCTRERISMHQDTTGACLLIQPAKK 182
QY 1239 SDAGWYTLAKNEAGIVSCTARLDIYAQWHQIPPPMSVRPSGRYSGLTSGKGLDIFS 1298
Db :|||||
183 SDAGWYTLAKNEAGIVSCTARLDIYAQWHQIPPPMSVRPSGRYSGLTSGKGLDIFS 242
QY 1299 SSMESTMVYSCSSRSRVESDEL 1320
Db :|||||

Db 243 SMESTMVCSSRSRVESDEL 264
RESULT 6
AAU17989
ID AAU17989 standard; Protein; 261 AA.
AC
XX AAU17989;
DT
XX
XX
DE 07-NOV-2001 (first entry)
XX Human immunoglobulin polypeptide SEQ ID No 134.
KW Immunoglobulin; signal transduction pathway protein; cancer;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
XX
PN WO200155315-A2.
XX
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01326.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

```
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-457725/49.
XX N-PSDB; AAS28777.
XX
XX Isolated novel immunoglobulin polypeptide for monitoring the presence
XX and progression of diseases and for diagnosis -
XX
XX Claim 11; SEQ ID No 134; 551pp; English.
XX
XX Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the
XX invention. The polypeptides and their associated polynucleotides can be
XX used to diagnose a pathological condition or a susceptibility to a
XX pathological condition in a subject by determining the presence or
XX absence of a mutation in a DNA sequence or determining the presence or
XX amount of expression of the protein. Alternatively the identification of
XX a binding partner to a sequence allows determination of changes in
XX protein activity. The sequences can be used as research tools for
XX receptors or other signal transduction pathway proteins that interact
XX with the polypeptides of the invention and can be used to treat, prevent
XX or diagnose various types of disorders such as neurological disorders,
XX cardiovascular disorders, gastrointestinal disorders, reproductive
XX disorders, immune system disorders, renal disorders, muscular disorders,
XX pulmonary disorders, proliferative disorders and cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 261 AA;
XX
XX Query Match 20.0%; Score 1380; DB 22; Length 261;
XX Best Local Similarity 99.2%; Pred. No. 7.3e-75;
XX Matches 239; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1060 ERKEPQERFFRPHFLQAPGDMVAHGRCLRDCKYSGLPPELTWLLNGQVLPDASH 1119
XX |||||
XX Db 1 ERKEPQERFFRPHFLQAPGDMVAHGRCLRDCKYSGLPPELTWLLNGQVLPDASH 60
XX
XX QY 1120 KMLVRETVGHSLLDPLDTQDAGTYKCIATNKGQNSFSLSVAKVKKAPVILEKLQ 1179
XX |||||
XX Db 61 KMLVRETVGHSLLDPLDTQDAGTYKCIATNKGQNSFSLSVAKVKKAPVILEKLQ 120
XX
XX QY 1180 NCGVPEGHVRLRCRVIGMPPPVYWKNDNETIPCTRERISMHDGTGYACLLIQPAKKS 1239
XX |||||
XX Db 121 NCGVPEGHVRLRCRVIGMPPPVYWKNDNETIPCTRERISMHDGTGYACLLIQPAKX 180
XX
XX QY 1240 DAGWYTLTSAKNAAGIVSCTARLDIYAQWHHQIPPPMSVRPSGSGRYSLTSKGLDIFSASF 1299
XX |||||
XX Db 181 DAGWYTLTSAKNAAGIVSCTARLDIYAQWHHQIPPPMSVRPSGSGRYSLTSKGLDIFSASF 240
XX
XX QY 1300 SMESTWYVSCSSRSRVESDEL 1320
XX |||||
XX Db 241 SMESTWYVSCSSRSRVESDEL 261
```

RESULT 7

```
AAU05396 *
ID AAU05396 standard; Protein; 26926 AA.
XX
XX AAU05396;
AC
XX
XX 24-OCT-2001 (first entry)
DT
XX
XX Human titin (connectin) protein sequence.
DE
XX
XX Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
XX titin-related disease; zebrafish; heart failure; heart disease.
XX
XX Homo sapiens.
OS
XX
XX WO200151666-A1.
PN
XX
XX 19-JUL-2001.
XX
XX 12-JAN-2001; 2001WO-US01212.
XX
XX 12-JAN-2000; 2000US-0175787.
XX
XX (GEO ) GEN HOSPITAL CORP.
XX
XX Fishman MC;
XX
XX WPI; 2001-451869/48.
XX N-PSDB; AAS05390.
XX
XX Determining if a subject has or is at risk of developing a
XX titin-related disease or condition, particularly heart failures,
XX comprises detecting the presence of a mutation in the titin gene -
XX
XX Disclosure; Page 57-111; 114pp; English.
XX
XX The present sequence representing human titin (also known as connectin)
XX is described in an invention relating to a novel method for determining
XX whether a subject has or is at risk of developing a titin-related
XX disease or condition. The method comprises analysing a nucleic acid
XX sample from the subject and detecting the presence of a mutation
XX (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
XX titin gene which indicates that the subject has or is at risk of
XX developing a titin-related disease. The zebrafish which has a phenotype
XX similar to mammalian heart failure is used as a model. The method is
XX useful for detecting an increased likelihood of heart disease, such as
XX heart failure, in a patient, so that appropriate intervention can be
XX instituted before any symptoms occur. The method may also be used to
XX facilitate determination of etiology of an existing heart condition,
XX such as heart failure, to identify compounds that can be used to treat
XX or prevent heart conditions, in prenatal genetic screening, e.g. to
XX identify parents who may be carriers of a recessive titin mutation.
XX Compounds identified using the methods may be used to treat patients
XX that have or are at risk of developing heart disease, e.g. heart
XX failure.
XX
XX Sequence 26926 AA;
XX
XX Query Match 11.6%; Score 797.5; DB 22; Length 26926;
XX Best Local Similarity 20.4%; Pred. No. 2.7e-37;
XX Matches 331; Conservative 205; Mismatches 489; Indels 599; Gaps 41;
XX
XX QY 198 SFDLSRRERRSSVIPIPADTRDNEVNHAEQOEAKRREAQAAASGAAGDTTPGSSPS 257
XX |||||
XX Db 57 SFSFD-----GRAKLTIPTAVTKANSGRYSL-----KATNGSGQATSTAE 94
XX
XX QY 258 SLYEPEPLGQPPRTQKLRSREVPEGTRVQLDCIVGIPPPQVRYWYCEKLENSPDHI 317
XX |||||
XX Db 95 LLVKAET--APPNFVORLQSMTRVQSGVRLQVRVTGIPTPVWKFYRDGAIEIQSLDFQI 152
XX
XX QY 318 VQAGNLHSLTIAEAFEDTGRYSCFASNIYGTDTSTSAFIYIEG-----VSSSDSEGD 369
```


PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0198874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457725/49.

N-PSDB; AAS28841.

Isolated novel immunoglobulin polypeptide for monitoring the presence and progression of diseases and for diagnosis -

Claim 11; SEQ ID NO 198; 551pp; English.

XX Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the
CC invention. The polypeptides and their associated polynucleotides can be
CC used to diagnose a pathological condition or a susceptibility to a
CC pathological condition in a subject by determining the presence or
CC absence of a mutation in a DNA sequence or determining the presence or
CC amount of expression of the protein. Alternatively the identification of
CC a binding partner to a sequence allows determination of changes in
CC protein activity. The sequences can be used as research tools for
CC receptors or other signal transduction pathway proteins that interact
CC with the polypeptides of the invention and can be used to treat, prevent
CC or diagnose various types of disorders such as neurological disorders,
CC cardiovascular disorders, gastrointestinal disorders, reproductive
CC disorders, immune system disorders, renal disorders, muscular disorders,
CC pulmonary disorders, proliferative disorders and cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 159 AA;

Query Match 10.5%; Score 724.5; DB 22; Length 159;
Best Local Similarity 89.68; Pred. No. 6.9e-36;
Matches 138; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

QY 1142 GTYKCIATNKTQNSFSLSVAKVKKAPVILEKLCQNGVPEGHVRLRCRVIGMPPP 1201

Db 8 GTYKCIATNKTQNSFSLSVAKVKKAPVILEKLCQNGVPEGHVRLRCRVIGMPPP 67

QY 1202 VFYWKNDNETPCTRERISMHODTGYACLLIQPAKKSADAGWYLSAKNEAGIVSCTARL 1261

Db 68 VFYWKNDNETPCTRERISMHODTGYACLLIQPAKKSADAGWYLSAKNEAGIXSCTARL 127

QY 1262 DIYAQWHHQIIPP--MSVPSGSRVSGSLTSKGLDI 1294

Db 128 DIYAQWHHQIIPXCLPAQXQSRISN--SKGLDI 159

RESULT 11

AAB40875

ID AAB40875 standard; Protein; 130 AA.

XX AAB40875;

AC AAB40875;

DT 08-FEB-2001 (first entry)

XX Human OREX ORF639 polypeptide sequence SEQ ID NO:1278.

DE Human OREX ORF639 polypeptide sequence SEQ ID NO:1278.

XX Human; open reading frame; OREX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antilarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC75084.

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 1115-1116; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
CC which represent the human OREX open reading frames 1 to 3161. The OREX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antilarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an OREX-associated disorder. The
CC nucleic acids can be used to express OREX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 130 AA;

Query Match 9.5%; Score 656; DB 21; Length 130;

Best Local Similarity 98.5%; Pred. No. 6.7e-32;

Matches 128; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 806 QFQPRCVSPVPVPTSRIONPVAFSLVPSLPAIPPTNAMKLPRSPAPMSQGLAKNT 865

Db 1 QFQPRCVSPVPVPTSRIONPVAFSLVPSLPAIPPTNAMKLPRSPAPMSQGLAKNT 60

QY 866 KSPQPVNDNIRETKNAVIRDLGKKITFSDVRPNQOEYKISSFFQRLMNEIEFRLETPV 925

Db 61 KSPQPVNDNIRETKNAVIRDLGKKITFSDVRPNQOEYKISSFFQRLMNEIEFRLETPV 120

QY 926 DESDDEIQHD 935

Db 121 DESHDEIQHD 130

RESULT 12

AAU84351

ID AAU84351 standard; Protein; 1953 AA.

XX AAU84351;

XX AAU84351;

DT 08-MAY-2002 (first entry)

XX Protein MYLK differentially expressed in breast cancer tissue.

DE Human; diagnosis of breast cancer; endometrial cancer; breast tumour;

XX MAI; mitotic activity index; cytostatic.

XX Homo sapiens.

XX WO200210436-A2.

PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-658660/75.
DR N-PSDB; ABL02768.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 2787; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4796 AA;

Query Match 7.9%; Score 545.5; DB 22; Length 4796;
Best Local Similarity 21.1%; Pred. No. 3.4e-23;
Matches 257; Conservative 170; Mismatches 484; Indels 305; Gaps 42;

Qy 215 IPADTR-DNEVNHAEQAEKREAEQAASEAGDTPGSSPSLYVEEPLGPPRETQ 273
Db LPSNFTGTESLQLEETHMKREE-----LVTEDSQNPDKPTE 3680

Qy 274 KLRSR-EVPEGTREVLDGIV--VGIPPPQVRWYCEGKELENSPDTHIVQAGNLSLTAIE 330
Db EIKNDLVPEGGPIHFDCEVPEVGPDPTRIEWFYNGHWATGSRVHQLNDFGFALDVIDY 3740

Qy 331 AFEEDTGRYSCFASNIYTDTSABIYIEG----VSSSDSGDPNKEEMNRIQKPNVSS 386
Db IYARDSGEYTCRATNKGWTATTSARVTCCKHNIYVESQLPEGMTSEKLKELEGR---- 3796

Qy 387 PPTTSVIPPVAPVQAHLVAPRVATIQCCSPNTYLOGLDGKPIIAAPVETKMLNLSA 446
Db IPEAPKVVEE-----VFGPPKFTTQITSVTV 3822

Qy 447 SEGQLVVFECRY--KGAPSPKVWYREGTLIEDSPDFRILQKKPRMAEPPEICTLVIAE 504
Db DEAEAVRECEQVEPKTDFSLRVWYRNGKPLPSGHRYNIFDM-----GFVSLDILY 3874

Qy 505 VPAESGCGFTCTASKNYG-----TVSS-----IAQLHV-RGNEDLSNNSLHSANSTT 551
Db VYGEDSGEYVCAINNYGDEDRATVSVCKLPTILLQNVPRG----MKRSDALTQMEATI 3931

Qy 552 NLAATEPQSPSPHSPSPSVEOPPK-----PKLEGVLVNHNEPRSS---- 591
Db KYTSEVHLTEDDLDPDRKQPPRVTVQIKQLTITEMAVTKFECQLAPVGDPNMKVWF 3991

Qy 592 -----SRIGLRVHFLPEDDKGS-----EASSEA 615
Db FNGKPLLHKNRFQPIYDFGYVAMNFGWVYPE-DSGEYVCRATNLYGDKETRAILKVSCKP 4050

Qy 616 GVWVTRQTRPSQXBRFNQATKTPPEPSFPVKEPPVLAQPKLSTQLQQLHNQVLLSQH 675
Db GIVYDSQL-PAHQMSIDIRIMEASQWVVPDEVPD--AKPRTRKVFVSKLEPQTVEE-- 4105

Qy 676 QLQNPPSPSPKEF-----PFKMTVLNSNAPPAVITTSKKQKAPSSQTFSLARPYF 726
Db -----GDPARFCVRVTGHPRPVMMVLNGH-----TVVHGSRYKLTNDGMFHLDPVKTR 4154
Homo sapiens.

Qy 727 FPST-----NTAATVAPSSSPVF-----TLSSTPQ-----T 753
Db QYDTGKVEVIARNVSGESIAATTELKVVARSDDYRNVLKNSPRWVDYELAAYQKEROENE 4214

Qy 754 IORTVSKRESLLVSHPSVOTKSPGGLSIQNEPLPGPTEPTPPFTFSPGNSQPPRCVS 813
Db LKVFDERKQVLSQSSTLTK--GV---EHLKPKQVKPPTPDQONVKA----- 4258

Qy 814 PIPVSPTSRIQNPVAFISSL--VLPSLPAIPPTNAXLP--RSAPSMPSQGLA----- 861
Db ----KKSDDYNNKQTLTETQQLKMETNLRDTHQVAPGEKVVSSSQAKMAQSYEENLQ 4314

Qy 862 -KNTKSPQPVNDNIRETKNAVIRDLKKITFSVDRNQOEYKISSPEQRLMNEIERL 920
Db EKTSTTEVQAAPPGIAQPSSESVHG-----REVHNKQ-----QQQKEIQDGL 4359

Qy 921 E--RTPVDESDETOHD-----EIPTKGCIAPIDPKRLKHFRTVTEGSPVTTCKIVG 970
Db EITRKITATETTEVEHKGCTIQERVVQGVKPAKAPVFTTKIQPCRVFENEQAKFEVEFEG 4419

Qy 971 IPVPKVVYFKDGKQIKSKNEHCMMRREGDGTCSLHIESTTSDDOGNYTIMAANPQGRISC 1030
Db EPNPTVKWYRESFPI--QNSPDQLIHFTSGKSIILIQVFVEDSAVFCVAENRGGTAKC 4477

Qy 1031 SGHLMVQSLPIRSRLTSAGQSHRGSRVQERDKPELQERFRPHFLOAPGDMVAHEGRLC 1090
Db SANLVVEE-----RRRACKGGIQP-----PSEVTTIQSTTATGQLA 4514

Qy 1091 RLDCKVSGLPPELPTWLLNGQPVLPDASHKMLVRETGVHSLLDLPLTQRDAGTKYCIATN 1150
Db RFDAKVTCTRLDVLVWLNKGMKIQPSIKFKML-BEDSVHTLLIIEPFAEDSGRYEAVN 4573

Qy 1151 KTGO-----NSFSLELSVVAKEVKAPVILEKQNCVPEGHPVRLRCRVIGMPPP 1201
Db AAGEARCDGDCIVOSPSPKPEPTTPGSEKAPHIVEQLKSQTVESGKSVIFRCRVDGRPTP 4633

Qy 1202 VFYWKK-DNETIPCTRERISMHQDITGACLLIQPAKSKDAGWYTLTSAKNAGIVSCTAR 1260
Db TARMWGENFVKPSRYFQMSRQGE--YYQLVISEAPEDESGYKCAENKLGSIQTSQAQ 4690

Qy 1261 LDYVAQNHQIPTPMS 1276
Db LKVRPIENLADPPTIT 4706

RESULT 14
ABP60991
ID ABP60991 standard; Protein; 5635 AA.
XX
AC ABP60991;
XX
XX
DT 10-SEP-2002 (first entry)
XX
DE Novel human protein. SEQ ID 78.
XX
KW Human; cytostatic; vulnery; antiarteriosclerotic; antiparkinsonian;
KW neurotropic; neuroprotective; immunosuppressive; haemostatic;
KW antiinflammatory; cardiact; antitumor; virucide; antithyroid;
KW cerebroprotective; anorectic; metabolic; vaccine; cancer; infection;
KW wound healing disorders; atherosclerosis; parkinson's disease;
KW Alzheimer's disease; autoimmune disorder; haematopoietic disorder;
KW inflammation; neoplastic disease; nervous system disorder;
KW cardiovascular disorders; pancreatitis; respiratory disorder;
KW hyperproliferation; systemic autoimmune disease; hyper-immunity;
KW developmental abnormality; gastrointestinal ulceration; neuropathy;
KW haematological disease; metabolic disease; sperm dysfunction;
KW thyroid disorder; hypothyroidism; brain damage; colitis;
KW cone photo- transduction deficiency; neurological disease; stroke;
KW angiogenesis; ovulation disorder; spinal cord; thyroid gland; heart;
KW trachea; thymus; lymph node; muscular system; obesity; anorexia;
KW growth abnormality; precocious puberty.
XX
OS Homo sapiens.

QY 1313 SVV 1315
: |
Db 2927 NNW 2929

Search completed: November 30, 2002, 19:04:50
Job time : 62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 19:03:55 ; Search time 19 Seconds
(without alignments)
2044.119 Million cell updates/sec

Title: US-09-818-990B-2

Perfect score: 6890

Sequence: 1 MODSIEASTSISQLRESY.....MESTMYSCSRVSDESL 1320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/1/iaa/PCU5_COMB.pep.*
- 6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	729	10.6	498	US-09-354-151-2	Sequence 2, Appli
2	335	4.9	277	US-09-354-151-3	Sequence 3, Appli
3	321.5	4.7	1297	US-09-540-245A-17	Sequence 17, Appl
4	318.5	4.6	1395	US-09-540-245A-15	Sequence 15, Appl
5	314	4.6	1651	US-09-540-245A-18	Sequence 18, Appl
6	311	4.5	601	US-08-795-868-16	Sequence 16, Appl
7	311	4.5	602	US-09-303-069-16	Sequence 16, Appl
8	311	4.5	602	US-09-134-250-16	Sequence 16, Appl
9	303	4.4	661	US-08-795-868-14	Sequence 14, Appl
10	303	4.4	661	US-09-303-069-14	Sequence 14, Appl
11	303	4.4	661	US-09-134-250-14	Sequence 14, Appl
12	300	4.4	1447	US-09-041-886-25	Sequence 25, Appl
13	300	4.4	1447	PCR-US94-05277-2	Sequence 2, Appli
14	267.5	3.9	607	US-08-752-307B-12	Sequence 12, Appl
15	267.5	3.9	607	US-09-707-802-12	Sequence 12, Appl
16	267.5	3.9	607	US-09-991-326-12	Sequence 12, Appl
17	263.5	3.8	612	US-08-752-307B-11	Sequence 11, Appl
18	263.5	3.8	612	US-09-707-802-11	Sequence 11, Appl
19	263.5	3.8	612	US-09-991-326-11	Sequence 11, Appl
20	263.5	3.8	1268	US-08-506-296B-28	Sequence 28, Appl
21	260.5	3.8	615	US-08-752-307B-9	Sequence 9, Appli
22	260.5	3.8	615	US-09-707-802-9	Sequence 9, Appli
23	260.5	3.8	615	US-09-991-326-9	Sequence 9, Appli
24	260.5	3.8	1381	US-09-540-245A-16	Sequence 16, Appl
25	257	3.7	380	US-09-877-730-4	Sequence 4, Appli
26	257	3.7	904	US-09-877-730-6	Sequence 6, Appli
27	257	3.7	985	US-09-877-730-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-354-151-2
; Sequence 2, Application US/09354151
; Patent No. 6458929
; GENERAL INFORMATION:
; APPLICANT: CARPEN, Olli
; APPLICANT: GRONHOLM, Mikaela
; APPLICANT: HEISKA, Leena
; APPLICANT: MYKKANEN, Olli-Matti
; APPLICANT: SALMIKANGAS, Paula
; TITLE OF INVENTION: Myocilin, A No. 6458929el Actin-Organizing Protein
; FILE REFERENCE: 0933-0142P
; CURRENT APPLICATION NUMBER: US/09/354,151
; CURRENT FILING DATE: 1999-07-16
; EARLIER APPLICATION NUMBER: 60/093,169
; EARLIER FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-354-151-2

Query Match	10.6%	Score 729;	DB 4;	Length 498;
Best Local Similarity	29.3%	Pred. No. 4.5e-41;		
Matches	185;	Conservative 101;	Mismatches 182;	Indels 164; Gaps 16;
QY	718	FSLARPKYFPFTNTTAATVAPSSPVFTLSSTPQTQRTV-----SKESLLV	765	
Db	2	FNVERPKHFIQSQNPCGSRQLQPGPETSFSFS--QTKQSSIIIPROQTEQRFSSATLS	59	
QY	766	SHPSVQTKSPGGLSIONEPLPGPTEPTPPPTFTSIFSGNQFQPCVSPVSPTSRTI--	823	
Db	60	SH-----ITMSSSAFPASPOQ-----HAGS-----NPGQRTVT	87	
QY	824	---QNPVAFSSVLPSPALPPTNAMXLPKSAPSMPSQGLAKK--NTKSPQPVN-----	872	
Db	88	TYNQSPASFLSSILPSQ---PDYNSSKIPSAMDSNYQSSAGQFINAKPSQTANAKPIPR	144	
QY	873	--DDNIRETKNAVTRDGLKKTFSVDRPNQOEYKISSFEQRLMNEIFRLERTPVDSDD	930	
Db	145	TPDHEIQGSKREALIQDLERLKKCD-----TLLHNGNQ	177	
QY	931	EIOHDEIPTGCKIAPIDKRLKHEFVTEGSPVTCTKIVGPVKVYWFKDGKISRNE	990	
Db	178	RLTYEERKMARLLGP-----ONAAVFAQ-----	202	
QY	991	HCKMRREGDGTCSLHIESTTSDDDGNTYIMAA NPQGRITSCSHLMVQSLPIRSRLTSAGQ	1050	

QY	124	SPTSSKESPOEAKRPQYCS-	-----TQSKVFLNKAADEFIELLSLFLKSHSKRI	173
Db	4714	SPASVDEAPQPSLPPAAQEGDLHLLWEALAKRRMSREPTLDSISELPE--EDGRSQR	4771	
QY	174	RPRACKNHKSLESQNKWOENSSFSDLSERRSSVPIPIPADTRDNEVNHAEQEA	233	
Db	4772	PQEA-----EEVAPDLSEGYSTADEL-----ARTGDADLSHTSSDDES	4809	
QY	234	KR-----REAPQAASEAAGDTPPGSSPSSLYEEP-----L	265	
Db	4810	RACTPSLVTYLLKAGRGPTSPLASKVGAPAAPSVKPQO--QOEPLAAVRPPLGLDSTKDL	4867	
QY	266	GQP-----PFTOKLRSREVPEGTRVOLDICVVGIPPP	298	
Db	4868	GDPMSDKAAVKIQAAFGYKVKEMKQOEGPMFSHTFGDTAQVGDALRLCEVCVYASKADV	4927	
QY	299	QVWYCEGKELENSPDIHVQAGN--LHSLTTAAEPEEDTGRYSCFASNIYGTDSIAEY	357	
Db	4928	RAWLKDGVGLDTRHHHLDQGDGTCSLLIAGLDADAGCYTCQVSNKFGQVTHSACV	4987	
QY	358	IEGVSSDSEGDGNKE-----EMNRIQKNEVSPPT--TSAVIPPAVPOA--	401	
Db	4988	VSG-SESEASSGGELDDAFRAARRLHRLFTKSPAEVSDDELFLSADSGPAEPEPA	5046	
QY	402	-----QHLYAOPRVATIQOCSPTNYQ-----GLDCK-----	429	
Db	5047	DMQTYREDEHFICIRFEALTEARQAVTRFOEMFATLGIVGRIKLVGEOGRPRVEMCISKET	5106	
QY	430	--PII-----AAPVFTKMLQNLASAGEOLVVFCEKRVKGAQSPKVEWYREGTFL	475	
Db	5107	PAPVPEPLPSLLTSDAAPVLTQLQNEQVQDGYPSFDCVVTGQPMPSRVWFKDKLL	5166	
QY	476	EDSPDFRILQKPRMAEPEICTLVIAEVFAEDSGCFTCTASNKYGTVSSIAQLHVRGN	535	
Db	5167	EEDDHVMINEDQOGH-----QLIITAVVPADMGVYKRLAENSMGVSSTKAELRV---	5216	
QY	536	EDLSNNGSLHSANSTTNLAAIEPQSPPHSEPPSVEQPKPKLECVLVNHNHNEPPSSSRIG	595	
Db	5217	-DUTSDTYDTAADATESSYFSAQ-----GYL-----SSR-----	5245	
QY	596	LRVHFNLPEDDKSEASSBAG-----VVTTQTRTPDSXOERFNGOATKTPES-EPVK	647	
Db	5246	-----EQEGESTTDEGQLPQWVEELRDLQVAPGTRLAKQLKVKGYPAERLWFK	5296	
QY	648	EPPLVIAKPLDSTQLQJLHNOVLL-----EQHQLQHPPPSSP	685	
Db	5297	DGQPLTASAHIRMTGKKILHTLEIISVTREDSGQYAAIYSNAMGAAYSSARLLVRGPDEP	5356	
QY	686	KEPFXMTVLNSNAPAVIT--SXQVKRAPSQTESL-----APKYYF	727	
Db	5357	EEKP-ASDVHEQLVPPMLERFTPKVKKGSSITFSVKEGVRPVTVHMLREAEARGVLW	5415	
QY	728	PSNTTAAATVAPS-----SPVFTLSSSPQTIQR	756	
Db	5416	IGPDTPGYTVASSAQOHSVLVDVGRHOGTYCTASNAAGALCSASLHVSGLPKYBEQ	5475	
QY	757	TVSKESLL-----VSHPSVQTKSPGGLSIQ--NEPLPPG-----	789	
Db	5476	EKVKEALISTFLOGTQTQAIASOGLTASPADJLGGORKEEPLAAKEALGHLSLAEVGTTEF	5535	
QY	790	-----TEPTPPPT--FSIPSGNQPPRCVPIVS-----PTSRIQ-----	824	
Db	5536	LQKLTQITDWSAKITQAKLQVPGDSDED---SKTSPASPRHGRSPSSISQESSSES	5592	
QY	825	-----NPVAPLSSVLP--SLPA	839	
Db	5593	EDGDARGEIFDIYVVTVADYLLPLGAEQDAITLREGQYVEVLDAAHPLRWLVKTKTSKSPS	5652	
QY	840	-----IPTNAMXLPRSAPSMPSOGLAKNKSQPQVNDNTRETKNAVIROLGKKITFSD	895	
Db	5653	ROGWWSPA---YLDRLKLSPEWGAEEAPPEFFGEAVSEDEYKARLSSVIOEL-----	5701	

Qy	896	VRPNQYKISSFEORLMNEIFP	-----RLRTPP-----	924	
Db	5702	-----LSSEQAFVEELOFLOSHHLOHLERCPHPVIAVAGOKAVIFNRVDRIGRFH	5751		
Qy	925	-----VDESDD-----	EQHDEIPTGKCIAPFQK-----	949	
Db	5752	SSFLQELQOCOTDDDDVAMCFIKNQAFAEQYLEFLVGRVQAESVVVSTAIQEFYKKAEEA	5811		
Qy	950	-----RLKIFRVTEGSPVTFCKIIVGIPVKVYWF-----	KDGKQISKRN-EHC	992	
Db	5812	LLAGDSPQPPPPPLQHY-----	LEQPVERVQRYQALLKELINKRNARNQNC	5857	
Qy	993	KMRREGDGTCG-----	LHIESTTSD-----	DGNVTMAANPQGRISCG	1032
Db	5858	ALLEQAVVVSALPQRAENKLHVSMLNYPGTLEALGEPHGFVWEGAPGARMWK	5917		
Qy	1033	H-----LMVQSLPIR-SRLTSAGQSHRGRSV-----	QERD-	1062	
Db	5918	HNRHVFILFRNHLVICKPRDRSRTDTVSYVFRNMMLSSIDLNDQVGGDRAFEVWQBER	5977		
Qy	1063	-----KEPLOER-----	FFRFHFLQAPGDVWAHEGRICRLDCKV	1096	
Db	5978	SVRYKYLLOARTAIKSSWKEICGIOORLALPWRRPPDEEBELADCTAELETVKLACRV	6037		
Qy	1097	SGLPPPELTWLLNQVLPDPASHKMLVRETGVHSLLDLPDRTORDAGTYKCIATNKTGONS	1156		
Db	6038	TGTPKPVISWYKDGKAVQVDPHILIEDPDGSCALILDSLTGVDSDGYWCFEAAASAG	6094		
Qy	1157	FSLELSVVAK-EVYKAPVILEKLQNGVPEGHFHVLECRVIGMPVPVFWKKDNETPCT	1215		
Db	6095	---NCSTLGLKILVQVPRPVNKVRASPFVEGEDAQFTCTIEGAPYQPTRWYKDGALLT	6151		
Qy	1216	RERISHQDITGYACILLIOPAKKSDAGWYTLTSAKNEAGIVSCTARLDI	1263		
Db	6152	NKFOITSEPRASGLLVIVRAAKSEDLGLYECELVNRIGLSARASAEIRI	6199		

RESULT 5

```

US-09-905-129-2
:
: Sequence 2, Application US/09905129
: Patent No. US20020137705A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Einat, et al
:
: TITLE OF INVENTION: GENES ASSOCIATED WITH
:
: TITLE OF INVENTION: AND USES THEREOF
:
: FILE REFERENCE: 540579-9-2007.2
:
: CURRENT APPLICATION NUMBER: US/09/905,129
:
: CURRENT FILING DATE: 2001-07-13
:
: PRIOR APPLICATION NUMBER: 09/802,318
:
: PRIOR FILING DATE: 2001-03-08
:
: PRIOR APPLICATION NUMBER: 60/207,821
:
: PRIOR FILING DATE: 2000-05-30
:
: PRIOR APPLICATION NUMBER: 60/084,944
:
: PRIOR FILING DATE: 1998-05-11
:
: PRIOR APPLICATION NUMBER: 60/085,673
:
: PRIOR FILING DATE: 1998-05-15
:
: NUMBER OF SEQ ID NOS: 25
:
: SOFTWARE: PatentIn version 3.0
:
: SEQ ID NO 2
:
: LENGTH: 2597
:
: TYPE: PRT
:
: ORGANISM: rattus species
:
: FEATURE:
:
: NAME/KEY: misc_feature
:
: LOCATION: (1)..(2597)
:
: OTHER INFORMATION: 'x' can be any amino
US-09-905-129-2

```

Query Match 6.0%; Score 416.5; DB 10; Length 2597;
Best Local Similarity 19.1%; Pred. No. 2.2e-12;
Matches 354; Conservative 216; Mismatches 552; Indels 733; Gaps 83;
Qy 2 QDDTEASTSTISQLLR---ESYLATRIHRGNNER---SRAEPSSNPFHGSPSGAAEGG 55

; LENGTH: 2597
; TYPE: PRT
; ORGANISM: Rattus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2597)
; OTHER INFORMATION: 'x' can be any amino acid
; -09-905-129-10

Query Match 6.0%; Score 416.5; DB 10; Length 2597;
Best Local Similarity 19.1%; Pred. No. 2.2e-12;
Matches 354; Conservative 216; Mismatches 552; Indels 733; Gaps 83;

QY 2 QDSIEASTSISQLR---ESYLAETRRHGNR---SRAEPSNPFCHFGSPGSAAGGG 55
Db 296 QEDNGSASTSQDFIEFGSLSLNMWXXSGNKADMVCSIQKPSRT-----SPTAFTE--- 347

QY 56 QDDLPDLAFLSQEELDESNIARLAINDYDLEKADETQARKRLSPDKMKHSNLSFEP 115
Db 348 -ENDYIMNASF-----STNLV-CSVDYNHIQPVWQLLALYSDSPILILERKPQLT--- 395

QY 116 NFOCDNPRSPTSKESQEAQRPOYCSSET-----QSKVP--LNKAADFIEEL 161
Db 396 -----ETPSLSRYKQVALRPEDITSIEADVRADFWFQOEKIVQLNRTATTSLT 448

QY 162 SSFLKSHSRKIRPRACKNHKSKLESQNKVMQENSSFSOLSERRERS-----SVP 212
Db 449 QIQFSTDAQIAL-PR-EMRAERLKWMTILMNN-----PKLERTVLVGGTIALSCP 498

QY 213 -----IPIPADTRNEV-----NHAEQOEAKRREA---EQAASBAAG 247
Db 499 GKGDPSPLEWLLADGSKVRAPYVSEGRILIDKNGKLEQMAFSDFAGLYHCISTNDAD 558

QY 248 GDITPGSSPSLXYE---EPLGQPPRTQKLRSEYPEGTRVQLDCIVGIPPPQVRWY 303
Db 559 ADV-----LTRYITVPEYGESTHDSGV--QHTVVTGETLDLCLSTGVDPASISWI 608

QY 304 CEKKELENSP--DTHIVQAGNLHSLTIAEAFEEDGRYSCFASNIYTDSTSAIYI--- 358
Db 609 LPGNTVFSQPSRQILNNGTLRILOYT---PKDQGHYQCVAA NPSCADPSFKVSVQKK 665

QY 359 -EGVSSSDSE-----GDPKNEENRQKNEVSPPTTSVAVIPAVP----- 399
Db 666 QORMVEHDREAGGGLGEPNSSVSLKQPSLKSASALTGTSEAGKQVSGVHRKKNHRDLI 725

QY 400 -----QAQHLVAOPRVATIOCCQ--SPNYLQGLDGKPI-IAAP 435
Db 726 HRRRGDSTLRFRHRRQLPLSARRIDPQWAALEKAKNSVPKKQENTTVKPVPLAVP 785

QY 436 VFTKMLQNLAS-----EGQLVPECEKVGAP--SPKVE-----WYREGTLI--- 475
Db 786 LVELTDEKDGASGIMPDEEFMVLKTRASGVGRSPADSGPVNHGFMTSIASCTEYSTV 845

QY 476 -----EDSPDFR-----ILQKKPRMAEPEICT---LVTAEPFAE--DS--- 510
Db 846 NPQTLQSEHLPDFEKLFSVTNGTAVTKWMNPSIASKIEDTTNQNPILIIIFPSVAEIRDSQA 905

QY 511 -----GCFTCTAS-----NKYGTVSSIAQLHVRGNE 536
Db 906 GRASSQSAHPVTGGMATYGHNTYSSFTSKASTVLOPINTESYGPQIPITGVSRPSSS 965

QY 537 DLSNNG-----SLRASNSTTNLAIEPQP-----SPPH 564
Db 966 DISSHTTADPSFSSHPSCSHTTASSLPHIPNNNTGNFPLSRHLGRERTIWSRGRVKNPH 1025

QY 565 SEP-----PSVEQP----- 573
Db 1026 RTPVLRHRRHRTVPAIKGPANKNVSOVPATEYPCGMCHTCPSAEGLTAVATAALSVPPSSH 1085

QY 574 ---PKPKLEGVLV----- 594
Db 1086 SALPKTNNGVIAEESTTVVKKPILLFKDKONVDIEIITTTKYSGGSENHVIPTASMT 1145

QY 595 GLRVHFNL---PEDKKGSEASSEAGVVTTTQTRDPDSXQERFNGQATKTPPSEFPKPP- 650
Db 1146 SAPTSVSLGKSPVDNSG-----HLSMPGTIQTGKDSVE-----TTPLPSP-PLSTPSI 1191

QY 651 PVLAKPKLDSTQLOQL--HNQ-----VLEQHQ--QNPDPSSPK-----EFPFXM 692
Db 1192 PTSTKFSKRKTPLHQIEFVNNQKKEGMLKNPYQFGLQKNPAKPLKPIAPLLPTGSSPSDS 1251

QY 693 TVLNSNAPPATWS-----XKOVKAPSSQTFSLARPKYFFPST-----NTT 733
Db 1252 TTTLSPPPALSALSTMAATQNGTEVSGARSLSAGKKQKPTNNSP--VLPSTISKRNTL 1309

QY 734 --AATVAPS--SSPVFTLSSPTQTIQTVSKES----- 762
Db 1310 NFLSTETPTVTSPTATASVIMSETQTRSRKAKDOIQKPRKNRNNANTTPROVSGYSAYS 1369

QY 763 -----LLVSHPSVQ-----T 772
Db 1370 ALTATDTPLAFSHSPRODDGNGNSAVAYHSTTSLTALTELFEKYTOTLGNTTALETLLS 1429

QY 773 KSPGGLSIGN-----EPL-----PPGTEPTPPPTFTSIPSGNQ---FQPRCVSPIPVS 818
Db 1430 KQESTTVKRASDTPPPLSSGAPPVET-PSPPPTKGVWTDVSKVTSFAQMTSRNVVTIY 1488

QY 819 PTSR---IQNPVA-----FLSVLPSPALPPTNAMXLPSPASPM 855
Db 1489 ESSRHNTDLQOPSAAEASPNPEIITGTTDPSNLFPSTVSPALRVYDKQNSKWKP----- 1542

QY 856 PSQGLAKNTKSPQPVQNDNIRETKNAVIRDLGKKITFSDVRPNQOEKYSFQRLMNE 915
Db 1543 -----SMP-----EHKYQLKSYSETIEKG 1562

QY 916 IEFRLERTPVDESDEIQHDEIPTGKCIA-----PIFDKR----- 950
Db 1563 KRPAVMSPP-----HLSLEASTHSHWNTQKAESVFDKPKQNPQNTSKHLPYVS 1613

QY 951 -----LKHFRVTEGSPVFTT-----CKIVGIQVPKYWFK--DGKQISKRNERCK 993
Db 1614 LPKTLALKPRIIGKKAASFTVPANSDVFLPCEAVGDPPLPIIHWTRVSSGXEISOGTQKSR 1673

QY 994 MRREGDGTCSLHIESTSDDDGNTTMAANPQGRISCSGHLWQSLPIRSLTSAGQSHR 1053
Db 1674 FHVLNPGT--LSIQRVSIQDRGOYLCSAFNPLGVDFHVSLSVVFYP----- 1718

QY 1054 GRSVQRDRKEPLOERFERPHFLQAPGDMVAHEGRCLRCLCKVSLGSLPELTWLLNGOPV 1113
Db 1719 --ARILDRHVK-----EITVHFGSTVELKCRVGMPPRTVSWILANQTV 1760

QY 1114 LPD---ASHKMLVRETGVHSLIDPLTQORDAGTKCIATNKTGNSFSELSVVAKEVKK 1170
Db 1761 VSETAKSRKVMWVTPDG--TLIIYNLSLYDRGFYKCVASNPSGDSLLVKIQVIT---A 1814

QY 1171 APVILEKLONC--GVPEGHVPRLECRVIGMPPPVFYKKDNET---IPCPRERISMHD 1224
Db 1815 PPVIEQKRQAIVGV-LGGSLLKPLCTAKGTQPPSVHVVLYDGTTELKPLQLTHSRFFLYPN 1873

QY 1225 TTGVACLLIOPAKKSDAGWYTLSSAKNEAG----- 1253
Db 1874 GT----LYIRSIAPSVRGTYECIATSSGSSERRVILTVGEGETIPRIETASQKWTEVNL 1929

QY 1254 ----IVSCTARLD-----IYAQWHHQIPPPMSPVSRPSGR--YGLSLSK 1290
Db 1930 GEKLLNCSATGDKPKRIINWRLPSKAVIDQW-HRMGSRHIVYPNGSLVGVGVTEK 1983

RESULT 7

US-09-905-129-13
; Sequence 13, Application us/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS TH
; TITLE OF INVENTION: AND USES THEREOF

FILE REFERENCE: 540579-2007.2
CURRENT APPLICATION NUMBER: US/09/905.129
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/802.318
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/207.821
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/084.944
PRIOR FILING DATE: 1998-05-11
PRIOR APPLICATION NUMBER: 60/085.673
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn version 3.0
SEQ ID NO 13

TYPE: PRT
ORGANISM: Rattus sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(2597)
OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-13

Query Match 6.0%; Score 416.5; DB 10; Length 2597;
Best Local Similarity 19.1%; Pred. No. 2.2e-12;
Matches 354; Conservative 216; Mismatches 552; Indels 733; Gaps 83;

Qy	2	QDDSTEASTSISQLLR---ESVLAETRIRGNNER---SRAEPSSNPCHFGSPSGNAEGG	55
Db	296	QEDNGASTSPQDFTEPPGSLNNTXXSGNKADMYCISQKPSRT-----SPTAFTE---	347
Qy	56	QDDLDLSAFLSQBELDESINLARLAINYPLEKADETOARKRLSPDQMKHSPNLSFEP	115
Db	348	---ENDYIMLNASF-----SINLV-CSVDYNIHQVPWQALLALYSDSPILLERKPLQT---	395
Qy	116	NFCQDNRPSPSKESQEAQRQVCST-----QSKVF--LNAADFIHEL	161
Db	396	-----ETPSLSRYKQVALRPEDIFTSIEADVRADPFWFQEQKIVLQLNRTATTLSTL	448
Qy	162	SSLEKSHSKRIRPRACKNHSKLESQNKVQENSSFSDLSEKERS-----SVP	212
Db	449	QIQFSTDAQIAL-PRA-EMRAERLKWTLMLMNN-----PKLERIVLVGGTIALSCP	498
Qy	213	-----IPIPADTRDNEV-----NHALEQQEAKRREA---EQARASEAG	247
Db	499	KGKDPSPHLEMLADGSKVRAPYVSEEDGRILDKNGKLEQMAQDSFDAGLYHCISTNDAD	558
Qy	248	GDTTPGSSPSLYB-----EPLGQPPRTQKLRSEVPEGTQVLDCTIVGIPPPQVRWY	303
Db	559	ADV-----LTYRITVPEYGESTHDSGV--QHTVVTGETLDLPCLSTGVPDASISWI	608
Qy	304	CEGKELENSP--DIHIVQAGNLHSLTIAEAFEDTGRYSCFASNIYGTGSTSAEYI---	358
Db	609	LPGNTVFSQPSRDRIILNGLTRILQVT---PKDGGHYQCAANPSGADFSFKVSVQKK	665
Qy	359	-EGVSSDSE-----GDPNKEEMRIQKPNREVSPPTTSVAPPVAPV-----	399
Db	666	GORMVEHDEAGGSLGEPNPSVSLKQASLKLASALTGSEAGKQVSGVHRKNNKRDLI	725
Qy	400	-----QAHLVQAPRVATTQQCQ---SPTNYLOGLDGKPT-IAAP	435
Db	726	HRRRGDSTLRRFRHRRLQPLSARIDPQRAWAALEKAKNSVPKKQENTTVKVPPLAVP	785
Qy	436	VFTKMLQNLSA---EGQLVVEFCRVKGAP--SPKVE-----WYREGTLI---	475
Db	786	LVELTDEKDKASGMTPPDEEFVWLTKKASGVGRSPTADSGPVNIGFMTSTASGTEVSTV	845
Qy	476	-----EDSPDR-----ILQKPRSMAPPEICT-----LVIAEVFAE--DS---	510
Db	846	NPQTLQSEHLPDFKLFVNTGNTAVTKSMNPSIAKIEDTTQNPIIIFPSVAEIRDSQA	905
Qy	511	-----GCTCTAS-----NKYGTVSSTAQLHVRNE	536
Db	1929	-----LYIRIAPSVRGTYECIATSSSGSERRVVILTVERGETIPRIETASQKTEVNL	1929
Db	906	GRASSQSAHPVTGGMATYGHNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSSS	965
Qy	537	DLSNNG-----SLHSANSTTNLAAIEQP-----SPPH	564
Db	966	DISSHTADPSFSSHPSGSHTASSLPHIPRNNTNGNPLSRHLGRERTIWSRGVRKNPH	1025
Qy	565	SEP-----PSVEQP-----	573
Db	1026	RTPVLLRRHRRHRTVRPAIKGPANKNVSOVPATEYPCMCHTCPSAEGLTVATAALSVPSSSH	1085
Qy	574	---PKPKLEGVLV-----	594
Db	1086	SALPKTNVGVIAEESTTVVKKPLLFKDKQNDIEIITTTTKYSGGESNHNVIPTASMT	1145
Qy	595	GLRVHFNL---PEDDKGSEASSEAGVVTROTTPDSXQERENGQATKTPESFPVKPEP-	650
Db	1146	SAPTSVSLGKSPVDNSG-----HLSMPTGIQTGKDSV-----TTPLES-PLSTPSI	1191
Qy	651	PVLAKPKLDSTLOQL--HNQ---VLLLEOHQL--QNPPSPSPK-----EFPFXM	692
Db	1192	PTSTKFSKRKTPLHQIFVNNQKKEGMLKNPYQFGLQKNPAKLPKIAPLLPTGOSSPDS	1251
Qy	693	TVLSNAPPVAVTTS-----XQVKAPSSQTFSLARPKYFFPST-----NTT	733
Db	1252	TTLTSPPALSTTMAATQNKGTVEVSGARLSAGKKQPTNSSP--VLPSTISKRSNTL	1309
Qy	734	-AATVAPS--SSPVFTLSSTPQTIQRTVSKES-----	762
Db	1310	NFLSTETPTVTSPTATASVIMSETQRTSKAKDOIKGPRKNRNANTTPROVSGYSAYS	1369
Qy	763	-----LLVSHPSVQ-----	772
Db	1370	ALTATDTPLAFSHSPRODGGNVSAVYHSTTSLLAITELFEKYVTQTLGNTTALETLLS	1429
Qy	773	KSPGGLSTON---EPL---PPGTEPTPPPTFTSIPSGNQ-----FQPRCVSPIPVS	818
Db	1430	KSQESTTVKRASTDPPLLSGAPVPT-PSPPPTKGVVDSKVTSFAFOTNRRVTIY	1488
Qy	819	PTSR---IQNPVA-----FLSSVLPSPALPPTNAMXLPSPAPSM	855
Db	1489	ESSRHNTDLOQPSAEASPNELIITGTDSPSNLFPSTVPAIRVVDKPKNSKWP-----	1542
Qy	856	PSQGLAKNNTKSPQVNDNIRETKNAVIRDLGKKITFSDVRPNQOEYKISFQORLME	915
Db	1543	-----SPWP-----EHYQLKSYSETIEKG	1562
Qy	916	IEFLRTPVDESDEIHOHEIPTGKICIA-----PIFDKR-----	950
Db	1563	KRPVMSMP-----HLSLPEASTHASHWNTQKHAESVFDKKPKQCNQNTSKHLPYVS	1613
Qy	951	-----LKHFRTVEGSPVFTT-----CKIVGIPVPKVYWFK--DGKQISKRNEHCK	993
Db	1614	LPKILLKPRIGGKAASFTVPANSDFVLPCEAVGDPLPIHWTYVSVSGXELISQGTQSR	1673
Qy	994	MRREGDGTCLHIESTTSDDDGNYTIMAANPQGRISCSGHLMVQSLPIRSRLTSAGQSHR	1053
Db	1674	FHVLPNGT--LSIQKVSIOQDQVQLCSAFNPLGVDHDFHVSLSVVFY-----	1718
Qy	1054	GRSVQERDKEPLQERFPHFLQAPGDMAHEGLCLDCKVSCGLPPELTWLLNGQPV	1113
Db	1719	-----ARILDHRVK-----EITVHFGSTVEKCRVEGMPRPTVSWILANQTV	1760
Qy	1114	LPD---ASHKMLVRETGVHSLLLIDPLTORDAGTYKCIATNKTGNSFSLSVLSVAKEVKK	1170
Db	1761	VSETAKSGKVVVTPDG--TLIIYNLSLYDRGFYKCVASNPSGQDLSLLVKLOVIT---	1814
Qy	1171	APVILEKQLQNC--GVPEGHVRLCEVRIGMPPPVFYWKNDNET-----IPCRTERRISMHD	1224
Db	1815	PPVITEQKQAIQVY-LGGSLLKPLCTAKGTQPSVHWVLYDGTTELKQLQLTHSRFFLYPN	1873
Qy	1225	TGVAACLIQPAKSKDAGWYTLAKNEAG-----	1253

; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 2587
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-905-129-16

Query Match 6.0%; Score 410; DB 10; Length 2587;

Best Local Similarity 19.4%; Pred. No. 4.4e-12;

Matches 283; Conservative 187; Mismatches 522; Indels 468; Gaps 59;

```
Qy 8 ASTSISQLRESYLAET--RRHGNERSRAEPSNP---CHFGSPSNAEGGGQDDLPD 62
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1325 ASVTYFQTERSQAQTQREQEPQKKNRTDPNISPDGSSGFTTPTAMT-----pp 1375

Qy 63 LSAFLSQBELDESVNLARLAINYDPLEKADQARKRLSPDQMKHSPNLSPEPNFCQDNP 122
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1376 ALATHSPENTTGISSITSPHSRLNLTVDIEELAQAOSTQTAKST--IASETTLSSKSH 1433

Qy 123 RSPSSKESPOEAKRPQVCSQTQS-----KKVFLNKA-----ADPIEELSFLFKSHS 169
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1434 QSTTTRKAS-LDTPIPLPSSSATLMPVPISPPTQRAVTDTRGDHSHFRMLTNTVVKLHE 1492

Qy 170 SKR-----IRPRACKN-----HKSLESQNKVMQENS-----SSPSDL 202
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1493 SSRINLQMPSSOLEPLTSTSNLHSTPMALTTVKSQNSKLTSPWAQYFQWHKPVSDI 1552

Qy 203 SERRERSVPPIPADTRDNEVNHAE-----QOEAKRREAE-QAASEAAGDGTTPGSSPS 257
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1553 AEKGKKKEV--SMLATGLSEATTLVSDWDQKNTKSKDFKPVQEAATTSKLLPFFSLS 1610

Qy 258 SLYEEPL---GQPRFTQKLRSEVEGTRVQDLCIVGIPPOVRW-YCEGKELENSP 313
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1611 RYIEKPRIVGGKAASET-----IPANSDAFLPCEAVGNPLPTIHWTRVSGLDLSRGN 1663

Qy 314 DIHIVQAGNLHSLTAFAFEEDTGRCFASNIYGTDSIAEIIYEGVSSSDSEGDPNKE 373
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1664 QNSRVQVLPNGTLSTQIRVEIODRGOYLCSASNLFGTDHLHVTLSV----- 1708

Qy 374 EMNRIQPNEVSSPPTTSVAPPVAPQAOHLVAQPRVATIQCCQSPNTYLGQDGKPIIA 433
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1709 -----VSYPFR----- 1714

Qy 434 APVTKMLQNLASGQGVLPCEKRVKGAAPKPVEM-YREGTLIEDSPDFRILQKKPRMA 492
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1715 --ILERRTKETVHSGSTVELKCAERGPSPTVTWILANQTWVSES-----SQG 1761

Qy 493 EPEEIC---TLVIAEVAEDSGCFTTASNKYGTVSSIAQLHVRGNEDLNNSGLHSAN 548
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1762 SRQAVTVVDGLVLHNLUSIYDRGPYKCVASNPGGODSLLVKIQV----- 1805

Qy 549 STTNLAATEPOPSPP-----HSEPPSVEQPPKPLEGLVLNVHNEPR-- 589
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1806 ---IAA-----PPVILEQRQVIVGTWGESLKLPTAKTPOPSVYVWLSDGTEVKPL 1855

Qy 590 -----SSRIGLRVHNLPEDDKGE---ASSEAG-----VVTTRQTRPDSXQER 631
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1856 QFTNSKLFNGTLYIR---NLASSDRGTVECIATSTSGSERRVWMLTMEERTVS--PR 1910

Qy 632 FNGQATKTPPEFPFPPPPVLAQPKLDSTQQLHQLHQLQLEQHLQNPSPSPKEFFFX 691
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1911 IEAASQKRTVNF-----GDKLLNCSATGEPRP----- 1939
```

```
Qy 692 MVLNSNAPPAVTTSTXKQV--KAPSSQTSFLARPKYFPFPSTNTTAATVAPSSSPVFTLSS 749
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1940 -----QIMWRLPSKAVVDGSGWIHYVPNGSLFSGVTEKDSGVYL--- 1979

Qy 750 TPQTIQRTVSKESLLVSHPSVOTKSPGGLSTQNEPLPPGPTTPTTTPFTSIPSGNQFQP 809
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1980 ---CVARNKMGDDLLMLHVSRLK-PAKIDHKQY-----FRKQVLHGKDFQV 2022

Qy 810 RCVSPIPVSPTRSQNPVAFSLSSVLPSPAIPPTNAMXLPSPASPMPSOGLAKKNTKSPQ 869
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2023 DCKAS-----GSPVPEISWSLPDGTMI--NNMQA----- 2050

Qy 870 PVNDNIRETKNAVIRDLGKKITESDVRPNOOEYKISSEFORLMEIEFRLETRTPVDES 929
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2051 ---DMSGHRTTTLFNNG-TLYENKVCV-AEEDGYTCYAQNTLGKDEMKVHLVITAAP 2105

Qy 930 DEIQHDEIPTGKCIAPIDFKRLKHFRVTEGSPVTTCKIVGIPVKYVW-FKDQKQISKR 988
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2106 RIRQSNK--TNK-----RIKAGDTAVLDCEVTGDPKPIFWLLPSNDMISFS 2150

Qy 989 NEHCKMRREGDGTCSLHTESTTSDDDGNYTIMAANPOGRISCSGHLMVQSLP----- 1040
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2151 IDRYTFHANG-----SLTINKVKLLDSGEYVCVARNPSGDDTKMYKLDVSKPPLINGLYT 2206

Qy 1041 -----IR-----SRLT----- 1046
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2207 NRTVIKATAVRHKKHFDCAEGTSPSEVMWIMPDNIFLTPAPYVGSRTVHKNGTLEIRN 2266

Qy 1047 -----SAGOSHGRS--RVOERDKEPLQERFTRPHFLQAPGDMVAHEGRLCLDC 1094
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2267 VRLSDSADFCVARNEGSESVLVQLEVMLEMLRPTFRNPENE---KIVAOLGKSTALNC 2323

Qy 1095 KVSGLPPELTWLL-----NGQVLPDASHKMLVRETGVHSLLDPLTQORDAGTKCI 1147
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2324 SVDGNPPEIITWILPNTGRFSNG-----POSQYLIASNG--SFLISKTTREDAGKYRCA 2376

Qy 1148 ATNKTGQNSFSELSLVAKEVKAPVILEKLQNC--GVPEGHVPVRLCVRIGMPPPVFYW 1205
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2377 ARNKVGY----IE-KLVILEIGQKPVILTYPAGTVKGI-SGESLSLHCVSDGIPKPNKW 2430

Qy 1206 KKDNETI---PCTRERISMHODTTGYACLLIQAPKAKSDAGWYTLISAKNEAGIVSCTARLD 1262
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2431 TMPSGYVVDPRQINGKYLHDNGT---LVIKEATAYDRGNVICKAONSVGHTLITVPVM 2486

Qy 1263 IYAQWHHQIPEPPMSVRPSGS 1282
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2487 IVA-----YPPRIITRNPERS 2501
```

Search completed: November 30, 2002, 19:11:34

Job time : 53 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 19:02:55 ; Search time 21 Seconds
(without alignments)
6042.735 Million cell updates/sec

Title: US-09-818-990b-2

Perfect score: 6890

Sequence: 1 MQDDSIESTSISQLRESY.....MESTWVYSCSSRSRVESDEL 1320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1666	24.2	772	T13078	KIAA0992 protein -
2	871.5	12.6	4162	T42633	connectin/titin -
3	801.5	11.6	26526	I38344	titin, cardiac mus
4	608	8.8	6642	T29757	protein UNC-89 - C
5	556.5	8.1	1906	I568235	myosin-light-chain
6	531.5	7.7	7962	I38346	elastic titin - hu
7	520.5	7.6	811	PN0689	connectin 1 - chic
8	513	7.4	2783	T34416	hypothetical prote
9	502	7.3	2541	T29340	hypothetical prote
10	468.5	6.8	5175	T20992	hypothetical prote
11	468.5	6.8	5198	T43290	hemiscetin precurs
12	429.5	6.2	1021	T42634	connectin/titin -
13	402	5.8	6831	A88852	protein unc-22 [im
14	402	5.8	6839	S57242	twitchin [similar]
15	402	5.8	7160	T27935	hypothetical prote
16	397.5	5.8	4391	A38096	perlecan precursor
17	375.5	5.4	3707	T18252	heparan sulfate pr
18	374	5.4	1323	PN0568	connectin 3B - chi
19	364	5.3	1176	TJ0583	myosin-light-chain
20	352.5	5.1	1147	AJ0587	myosin-light-chain
21	351.5	5.1	6658	T13931	projectin - fruit
22	329	4.8	1273	T42405	sax-3 protein - Ca
23	322	4.7	3375	T19821	hypothetical prote
24	321.5	4.7	423	T29549	hypothetical prote
25	319	4.6	1612	T30805	dufil protein - mo
26	317	4.6	1465	T43529	165K protein, skel
27	316.5	4.6	1344	T14316	rig-1 protein, mo
28	315.5	4.6	1651	T14160	transmembrane rece
29	315	4.6	1896	T08851	Down syndrome cell

30	307.5	4.5	1427	2	I51669	tumor suppressor -
31	301.5	4.4	1450	2	A44027	165K myofibrillar
32	300	4.4	940	2	A40985	projectin - fruit
33	300	4.4	1447	2	A54100	tumor suppressor p
34	299.5	4.3	1742	2	S24600	projectin - fruit
35	293.5	4.3	1535	2	S46224	peroxidase - fruit
36	288.5	4.2	2109	2	E89066	protein UNC-52 [im
37	288.5	4.2	2109	2	T33247	hypothetical prote
38	285	4.1	1240	2	T03097	CDO protein - huma
39	284	4.1	1488	2	T34418	hypothetical prote
40	282.5	4.1	1666	2	A48594	skelemin - mouse
41	281.5	4.1	1070	2	JC4593	protein-tyrosine k
42	277	4.0	2295	2	C88369	protein unc-52 [im
43	277	4.0	6805	2	S20901	titin - rabbit (fr
44	272	3.9	725	1	IJMSNG	neural cell adhesi
45	272	3.9	1115	1	IJMSNL	neural cell adhesi

ALIGNMENTS

RESULT 1

T13078

KIAA0992 protein - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13078, T08742

R:Nagase, T.; Ishikawa, K.; Suyama, M.; Kikuno, R.; Hirose, M.; Miyajima, N.; Tanaka

DNA Res. 6, 63-70, 1999

A:Title: Prediction of the coding sequences of unidentified human genes. XIII. The co

A:Reference number: Z17602; MUID:99246063; PMID:10231032

A:Accession: T13078

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-772 <NAG>

A:Cross-references: EMBL:AB023209; NID:g4589627; PIDN:BAA76836.1; PID:g4589628

A:Experimental source: brain

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z16471

A:Accession: T08742

A:Molecule type: mRNA

A:Residues: 628-772 <WAM>

A:Cross-references: EMBL:AL050093

A:Experimental source: adult uterus; clone DKF2p586L0518

C:Genetics:

A:Note: KIAA0992; DKF2p586L0518.1

Query Match 24.2%; Score 1666; DB 2; Length 772;
Best Local Similarity 45.2%; Pred. No. 1e-73;
Matches 361; Conservative 113; Mismatches 193; Indels 132; Gaps 20;

QY 621 RQTRPSXQBRFNGQATKTPSPSPVKEPPVLA-----KPKLDSTQLQQLH 667

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Db 7 RGIKPDT-----CPAPGRSPLQLP---LAPDAESGSGRRRGPDR-DPLKQLQL 53

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY 668 NQVLLSQHQ-LQNPDP---SSPKPEFP-----XMTVLNSNAPPAVT 704

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Db 54 NQRLQEGAGARQPPAPRSAPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 112

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY 705 TSXKQVKAPSSQSFSLARPKYFFPSNTTAAAT--VAPSSSPVFTLSSTPOTIORTVSKES 762

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Db 113 -----QSSGFNFARPKQFTAAQNLGPASGHGTPASSP---SSSLSPSPMFTPRQFG 162

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY 763 LLYSHFSVQ-----TKSPGGLSIQNEPLPP---GPEPTPPPTFTSIPSGNQFQPRCVS 813

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Db 163 RAPVPPFAQPFGAPEAPEAWGSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 217

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY 814 PIP-----VSPRSRI---ONPVAFSLVPSLPALPTTNAMXLPSPAPSMPSQGLAKKN 864

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Db 218 PSPGQASHCSSPATRFHGQGTAAFLSALLPSQPPPPAAVNALGLPK---GVTPAGFPKKA 274

| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY 865 TKSPQPVNDNIRETKNAVIRDLGKKTTFSDVRPN-----GVTGTFSDVRPN-----899

```
Db 275 SRTASDEIEQTKDAVIOGLERKLFKEDLLNNGQPRLTYSERMAARRLLGADSATVF 334
Qy 900 -----QOEYKISSFEORLMEIEFLRLERTPVDESDDIEOHDEIPTCKIAPIDFK 949
Db 335 NIQPEETANQETKVSCEORLISEIEYRLERSPVDESDEVOIGDVPVENGMAPFEM 394
Qy 950 RLKHFRVTEGSPVFTCKIVGIPVKVWFMDGKGQISKRNEHCKMRREGDGTCSLHTEST 1009
Db 395 KKKHYKIFEGMPVFTCRVAGNPKKIYWKDGKQISPKSDHYTIQRLDGTCSLHTTAS 454
Qy 1010 TSDDGNTYIMAAPOGRISCSGHLMVQSLPIRSLTSAGOSH-----RGRSRVQERDK 1063
Db 455 TLDDGNTYIMAAPOGRISCTGLMWQAVNQGRSPRSPSGCHPVRPRSRSDSGDEN 514
Qy 1064 EPLQERFRPHFLOAGDWAHEGRCLRLCKVSLGSLPPELTLLNGOPVLPDASHKMLV 1123
Db 515 EPIQERFRPHFLOAGDWTVOEGRCLRMDCVKSGLPTPDLWSLQDCKVPRPDSAHKMLV 574
Qy 1124 RETGVHSLIDPLTORAGTYKCIATNKTKGNSFSLSVVAKEVKKAPVILEKLQNGV 1183
Db 575 RENGVSIIIEPVTSRDAGIYTCIATNRAGNSFSELVVAKEAHKPPVPIEKLQNTGV 634
Qy 1184 PEGHPVRLCVRIGMPVFTWKNDNETIPCTRERISHQDQTTGYACLLIOPAKKSDAGW 1243
Db 635 ADGYPVRLCVRIGMPVFTWKNDNETIPCTRERISHQDQTTGYACLLIOPAKKSDAGW 694
Qy 1244 YTLAKNEAGVCTARLDIYAQMH--QIPPPMSVGRSGRYSLTSKGLDIFSASSM 1301
Db 695 YTVSAKNEAGVCTARLDVITQMHQSQSTKPKKVRPSASRYAALSQDGLDIKAAFP- 753
Qy 1302 ESTMVYSCSSRSVSEDEL 1320
Db 754 EANPSHLTLNLTALVESED 772

RESULT 2
T42633
connectin/titin - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42633
R:ajima, H.; Ohtsuka, H.; Kawamura, Y.; Kume, H.; Murayama, T.; Abe, H.; Kimura, S.; Ma
Biochem. Biophys. Res. Commun. 223, 160-164, 1996
A:Title: A 11.5-kb 5'-terminal cDNA sequence of chicken breast muscle connectin/titin re
A:Reference number: 222221; MUID:96254045; PMID:8660363
A:Accession: T42633
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4162 <YAJ>
A:Cross-references: EMBL:D83390; NID:g1513029; PIDN:BAAL1908.1; PID:g1513030
A:Experimental source: breast muscle
C:Keywords: skeletal muscle

Query Match
Best Local Similarity 23.5%; Score 871.5; DB 2; Length 4162;
Matches 327; Conservative 193; Mismatches 507; Indels 367; Gaps 43;

Qy 209 SSVPIPIADPRDNEVHA-----LEQOEAKREAEQAASEAAG-----DITPGSSPS 257
Db 30 SGFPVPEVSWYRGQVLSAATLPQVIFSFGRAKLVPISVTEANSRYTITQATNGSGQA 89
Qy 258 SLYVEEPL--GQPRFTQKLRSREVPETGRVQDLCIVVGPVPPQVRWYCEGKELENSPD 314
Db 90 TSTAELLVTAGTAPNFSORLOSMTARQGSQVRLDVRVTGTPVVKFYRQGVIEIQSPD 149
Qy 315 IHIVQANLHSLTTAEAFEDTGRYSFASNIYCTDSAEIYYIEGVSSSDSEGDPNKE- 373
Db 150 FOILQEGDLSLIAEAYPEDSGTYSYVSNATNVNNGRATSTAEILLIQ-----EEEAAPAKT 205
Qy 374 -----ENNRIOK-----PNEVSSPPTS 391
Db 206 KTIYSTAQISQTOARIEKKIETHFDARSLTSEWMTIEGAAAOQLPHKAPRMPPTSK 265
```

```
Qy 392 AVIPPAVPAQHLVAQPRVATIOQCQSPNTYNOGLDGKPIIAADPVFTKMLQNLSASEGOL 451
Db 266 SPTPEVITAKAQAQAQSPSPVRQSPSPVRHVRAPTPSPVRSV-----SPAGRI 314
Qy 452 VVFCRVKVGAPSP--KVE-----WYREG---TLIEDSPDFRI----- 483
Db 315 STSPIRVKSPSPIRKAQVVTPGAELVPPWRQEGYSATAEAKMKEVTSATEIRTEER 374
Qy 484 -----LQ-----KKPRMAEPPEICTLVIA-----EV 505
Db 375 WEGRYGLQEQVTTISGAAGEVAAGAKEVRKEPEKTPVPTVIIATDKAKEQERISTAREEI 434
Qy 506 FAEDSGCFTCTASKNG-----TVSSIAQLHVRGN----- 535
Db 435 SARHEQVHVSHQIEAKRAEAVATVVAADQAVRSPWETEQVDETYVVKKLTLEYGYKE 494
Qy 536 ---EDLSNNGSLHSANSTTNLAATIEPQSPSPHSEPPSV-----EOPPKPKLEGVL-VNH 585
Db 495 HAVKDEAQAQAEHHVATKEVKTVVVPEKHIPAAEKKEVHVSTEIKRETEAKIEKTIHIEH 554
Qy 586 NEPRSSSRIGLRVHFN-----LPEDDKGSASSEAGVVTTRQ-----TRPDSX 628
Db 555 PRPETAS-----PHFTVSKIAVAPKPDHTEVSIAGSAMATLEKELSATSAAQKITKPKVP 609
Qy 629 QERENGQATKTP-----PSFPVKKEPP-----PVLAKPKLDSTOLOLHNQVLLLEQHQLQ 678
Db 610 POLKPEHVKKIPESAPQPFTEAAETVKAHYDETKEVDVS-----IKGEAVREDHLLL 665
Qy 679 NPPSS-----PKEFPFXMTVL-----NSNAPPVATTSX 707
Db 666 RKESEAKVTETARVPVPAETPPTLVWGLKNKVTTEGESVTLCHISGHQPTVTWYR 725
Qy 708 KOYKAPSQTFSL-----ARPKYFPPTNTTAAATVAPSSSPVFTLSSTPQ 752
Db 726 EDYKIESMDFOITFKAGLARLVIREAFEDSGRFTCTATNKAGSVSTCHLVKVKSEET 785
Qy 753 TIQRTVSKESLLVSHSPSVQTK-----SPGGLSIQNEPLP-----GTEP 792
Db 786 ETRETISEKVTEKSYVETKDVVMDVSAAAEVSEGPVPPFPIRKPVVHKLIEGGS-- 843
Qy 793 TPPPFTSIFSGNQFQRCV---SPIVPSPTSR----- 822
Db 844 ---IIFECQVGGNPKPHVLMWKGKGVPLTTGYRYKSVKRYKRETKETGECKLEISMTFADDAEY 899
Qy 823 ---IQNPVAFLLSVLSPALPPIPNAMKLLPRSAPSMPSQGLAKKNTKSPQ-----PV-- 871
Db 900 TVIRNKFEASA---TVSULEEADYEAYIKSQOEMMYQTQVTAIYQVEPKVAEVAPPISY 956
Qy 872 -NDNINIRETKNAVIR-DLGGKKITFSDVRPNQOEYKISSFEORLMEIEFLRLERTPVD--- 926
Db 957 GDFDKEYEKEQALIRKKMAKDTVMVVRTFVEDEEFHISSEERLIKEIELRIKTLDELL 1016
Qy 927 ESDDEIOHDEIPTGKCIAPIDFKRLKHFRVTEGSPVFTCKIVGIPVKVYFWDGKQIS 986
Db 1017 EEDGEEMMIDISEAIGAGFDLRLKKNRYTFEGTGVTFHCKTGYPLPKIAWKDGKRI- 1075
Qy 987 KRNECHKRRREGDGTCSLHTESTSDDGNTYIMAAPOGRISCSGHLMVQSL----- 1039
Db 1076 RHGERYHMEVLQDGSASRLPVVLPEDEGIYTFASNMKNKAICSAKLYVEPVAPTATPG 1135
Qy 1040 -----PIRSRLTSAGQSHRGRSVQERDKEPLQERFRPHFLOAGPD 1081
Db 1136 YMPCPEVMRRYRSTSPRSRSPARSPSPAR-RLDETDEGOL-ERLKYPFVFLKPTS 1193
Qy 1082 MVAHEGRCLRLCKVSGLPPELTLLNGOPVLPDASHKMLVRETGVHSLIDPLTORDA 1141
Db 1194 VKSQGTARFDLKVVGPRMPETVWFHNGQVVDYTHKVIKEDGTQSLIIVPAMPEDS 1253
Qy 1142 GTYKCIATNKTKGNSFSLSVVAKEVKKAPVILEKLQNGVPEGHVPRLECRVIGMPPP 1201
Db 1254 GENAVIAQNAGRAGASVTLVSVEAKEDLVPRPREVERURNVSVKESRLHMAVKATGNPNP 1313
```


[illegible]

A;Residues: 1750-1906 <YOS>
A;Cross-references: EMBL:M96655; NID:g212744; PIDN:AAA49083.1; PID:g212745
A;Accession: S78216
A;Molecule type: DNA
A;Residues: 1750-1906 <VOW>
A;Cross-references: EMBL:M96987
R;Olson, N.J.; Pearson, R.B.; Needleman, D.S.; Hurwitz, M.Y.; Kemp, B.E.; Means, A.R.
Proc. Natl. Acad. Sci. U.S.A. 87, 2284-2288, 1990
A;Title: Regulatory and structural motifs of chicken gizzard myosin light chain kinas
A;Reference number: A35093; MUID:90192792; PMID:2315320
A;Accession: A35093
A>Status: preliminary
A;Molecule type: mRNA
A;Residues: 935-1438,'Q',1440-1906 <OLS>
A;Cross-references: GB:M31048; NID:g212660; PIDN:AAA49069.1; PID:g212661
R;Guerrero Jr., V.; Russo, M.A.; Olson, N.J.; Putkey, J.A.; Means, A.R.
Biochemistry 25, 8372-8381, 1986
A;Title: Domain organization of chicken gizzard myosin light chain kinase deduced fro
A;Reference number: A25810; MUID:87157587; PMID:3030394
A;Accession: A25810
A;Molecule type: mRNA
A;Residues: 1258-1438,'Q',1440-1906 <GUE>
C;Genetics:
C;Introns: 1735/3; 1779/1; 1819/1
C;Superfamily: myosin-light-chain kinase, nonmuscle; fibronectin type III repeat homo
C;Keywords: alternative initiators; ATP; calmodulin binding; phosphoprotein; phosphot
F;542-599/Domain: immunoglobulin homology <IMM1>
F;935-1906/Product: myosin-light-chain kinase, 108K, smooth muscle (from 5.5kb transcr
E;1098-1158/Domain: immunoglobulin homology <IMM2>
F;1451-1708/Domain: protein kinase homology <KIN>
F;1459-1467/Region: protein kinase ATP-binding motif
F;1750-1906/Product: telokin (kinase-related protein KRP) (from 2.7 kb transkript) #
F;1808-1869/Domain: immunoglobulin homology <IMM3>

Query Match 8.1%; Score 556.5; DB 1; Length 1906;
Best Local Similarity 21.9%; Pred. No. 2.9e-19;
Matches 237; Conservative 151; Mismatches 438; Indels 255; Gaps 36;

QY 267 QPFRFTKLRSVEPECTRVQLDCIVGIPPPVRWYCEGKELENSPDHIHVOAG--NLH 324
Db : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 EPAFTLPNRIRVOLGATARFEKGVRGYEPQITWRNGHPLEG-DHYVVDSIRGIF 84
QY 325 SLTIAFAEDTGRTGSFCASNIYGTDSTSAIYIEGVSSSDSEGDPNKEMRNTRKQNEV 384
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D8 SLVIKGVQEGSGRYTCEAANDGGVRQVTVELTVNEG-----NSLKYSLSL 129
QY 385 SSPPTTSNAVTPVPQAHLVAQPRVATIOQCQSPTYNQLOGDKPII---AAPVFTKML 441
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 130 SSAKTPGRL--SVPPVEH-----RPSIWGESPEPKFATKP 162
QY 442 QNLASRGQLVWFPCRVKGAPSPKVWREGLTEDSPDFRILOKKPRSAEPEICTLV 501
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 163 NRVVVRGQTGRCSCKITGRPOVTTK-----GDHLQONERFNPFKTGIOYLE 214
QY 502 IAEVFAEDSCGCTCTASNKYGTVSSIAQLHVRRGNEDLSNGLSHSANTTNLAIEPQPS 561
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 215 IQNVQLADAGIYCTVTVNSAGKASVSaelTVQG-----PDKT 251
QY 562 PPHEPSPVQPQPK-KLEGLVLNHNPRESSRIGLRVHF-----NLPDDKGSASEA 615
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 252 DTHAQP--LCMPKPPTTLATKAENSDFKOATSINKELKSTSTELMVETKDLRSAKK 309
QY 616 GVVTTRTPDSXGERENGQATKTPESFPVKPEPPPVLAKPLDSTOLQLHNOLLVLEQH 675
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 310 TFFTSRAK-DGKOGQ-NQEANVPLESRGTGPQV-----LQKTSITTILQAV 357
QY 676 QLQNPPSPSSKEPFFXTMTVLNSNAPPATTXXKQVKAPSSQTFSLARPKYFFPSTNTAA 735
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 358 KAQPEPKAEQO-----TTFIRQAE-----DRKRIVQLMTYTTT- 390
QY 736 TVAPSSSPVFTLSSTPOTIQRTVSKESLLVSHPSVQTKPGGL--STONEPLPPGPTEPT 793
Db ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
D6 391 ----QENPSLTGVQSPR-----SRETENRAGVKSVKEEKREP-----424

```
Qy 794 PPTTSIPSGNQPRCVSPIPVPSITSRIQNPVAFSLSSVLPSPALPPTNAMXLPSPAP 853
Db 425 -----LGIPP--QFESR-----POSLEASGEQEIKFKSKV-----SGKPRPDV 460

Qy 854 SMPSQGLAKNWKSPQPVNDNIR--ETKNNAVIRDLGKKITFSVDRNPQOQYKISSFEOR 911
Db 461 EMFKEGVPPIKTEGIGIQIYEEDGTHCLMKKACLDGSG-----SYSCAAFNPR 507

Qy 912 LMNEIERL--ERTPVDSDDDEIOHDEIPTGCKIAPTFDKRLKHFRVTEGSPVTFCKIV 969
Db 508 QGTSTWLLTVKRPKVEE-----VAPCFSSVLKGTCTVSEGDQFVLQCYVG 552

Qy 970 GIPVPKYVWFKDGKQIISKRNHCKMRREGDGTCSLHIESTTSDDDGNYTTMAANPOGRIS 1029
Db 553 GVPVPEITWLLNEQPIQYAHSTFEA---GVAKLTVDQDALPEDDGIYTCLEANNAGRAS 607

Qy 1030 CSGHLMVQSLPIRSLTSAGOSHGRSRVORDEKEPFRPHFLQAPGDMVAHGEGL 1089
Db 608 CSAQVTVK-----EKKSSKABGTQAQAKLNKTFAPIFLKLGLTDLKVMDSQ 653

Qy 1090 CRLDCKVSGLPPPELTWLLNGQVLPDASHKMLVRETGVHSLLDPLTQRDAGYKCIAT 1149
Db 654 VIMTVEVSANPCPEIILWHNGKEI-QETEDPHFEKKNEYSLYIOEVFPEDGTGKYTCBAW 712

Qy 1150 NKTQNSFSLSLSVAKREVKKAPVILEKLNQGVPEGHVPRLCVRVIGMPPVPVYWKKN 1209
Db 713 NELGETQATLTQVEPDQDGIQWFIKPSRVSVAAGONVLISCAIAGDPPTVHWFKDG 772

Qy 1210 ETI-PCRTERISMHDQDTGYACLIQIAPKASDAGWYTLISAKNEAGIVSC-----T 1258
Db 773 QEITPGTGCEILQND---IPTLLRNQSRHAGQYETQLRNQVGECSQVSLMLRESSA 829

Qy 1259 ARLDIYAQWHHQIIPPMVSVPSSGRSYSLASKGLDIFSAFSSMESTWYVSSCSRVSSED 1318
Db 830 SRAEMLRDGRES---ASSGERRDGNGYALT-----FGTSGGFKKS---SSETRAABEEQ 878

Qy 1319 E 1319
Db 879 E 879

RESULT 6
I38346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38346
R:Labait, S.; Koimerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38346
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: EMBL:X90569; MID:g1017426; PIDN:CAA62189.1; PID:g1017427
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 7.7%; Score 531.5; DB 2; Length 7962;
Best Local Similarity 21.3%; Pred. No. 2.9e-17;
Matches 256; Conservative 152; Mismatches 436; Indels 359; Gaps 44;

Qy 267 QPPRFTOKLRSREVEPTRVOLDICIVGIPPPQVRWYCEGKELENSDPIHIVQAGNLHSL 326
Db 3683 EPPYFIEPLEHVEAIVEGEPATLQKVDGTPETIRISWYKEHTKLRSAPAYKMQFNKNVASL 3742

Qy 327 TIAAFEDTGRYSCFASNIYGTSTSAEYIEGVSSSDSDPNKEEMNRIQKPNVESS 386
Db 3743 VINKVDHSDVGEYSCRADNSVGAVASAVLVI----- 3774
```

```
Qy 387 PPTTSVIPPAVQAOHLVAPQPRVATIQCCQSPNTYLOGLDGKPIIAAPVFTKMLQNSA 446
Db 3775 -----KARKL-----PPFARKLKDVEH 3792

Qy 447 SEGOLVVFECRVKAPSPKVEWYREGTLIEDSPDRILQKKPRSMARPEICTILVIAEVF 506
Db 3793 TLGFPVAFECRINGSLEPQVSWYKDGVLKDDANLQ-----TSFVHNVATLIQLQTD 3844

Qy 507 AEDSGCFTCTASNKYGTIVSSIAQL-----HVRG-- 534
Db 3845 QSHIGQYNCNSASPLGTASSAKLILSEHEVPPFDLKPVSVDLALGESGTFKCHVTGTA 3904

Qy 535 -----NEDLSNNGS--LHSANSTTNLAAIEPOPSPPH-----SEPP 568
Db 3905 PIKITWAKDNREIRPGNYKMTLVENTATLTVLKVGKGAGQYTCYASNTAGKDCSAQL 3964

Qy 569 SVEQPPK--PKLE--GVLVNHNH--PRSSSRIG----- 595
Db 3965 GVQPPRFIRKLEPSRIVKQDEFTRYECKIGGSPEIKVLMYKDETEQESSKFRMSFVDS 4024

Qy 596 ---LRVHFNLPEDDKGSEASSEAGVVTTRQTRPDSXQERFNGQATKTPEPSFPVKPEPPV 652
Db 4025 VAVLEMH--NLSVEDSG-----DYTCEAHNAAGSASSSTSLKYKEPPIF 4066

Qy 653 LAKP-KLDSTQLQOLHNOVLLEQHLQNPDP-----SSPK-----EPFEXM 692
Db 4067 RKRPHPETILKGDVHLEC-----ELQGTFFFVHSWKDKRELRSKKYKIMSENLTSI 4121

Qy 693 TVLNSNAPPVTTSSXKOVKAPSQSTFSLARPKYFFPSTNTTAATVAPSSSPVFT--LSST 750
Db 4122 HILNVD---AADIGEYQCKATND-----VGSDDTCVGSIALKAPPRFVKLSDI 4166

Qy 751 PQTQRTVSKESLLVSPSVQT---KSPGGLSITONEPLPPGTEPTPPPTTSPSQNQF 807
Db 4167 STVYGVKEVQLQTTIEGAEPISVVMFKDKGRIRESDNINWISYSENIATLOFSRVEPANAG 4226

Qy 808 QPRCVSPIPVPSPTSRIQNPVAFSLSSVLPSPALPPTNAMXLPSPAPSMPSOGLAKNWKTS 867
Db 4227 KYTC-----QIKND-AGMOECFATLSVLEPAITVEKPEIKVTGDTCTLECTVA 4275

Qy 868 PQPVNDNIRETKNNAVIRDLGKKITFSVDRNPQOQYKISSFEQ-----RLMNEIEFRLERT 923
Db 4276 GTP-----ELSTKWFKD-GKELT-----SDNKYKISPFNKVSGLKIIN----- 4312

Qy 924 PVDESDDDEIOHDEI--PTGK--CIA-----PFDKRLKHFRVTEGSPVTFCKIV 969
Db 4313 -VAPSDSGVYSFEQNPVGVKDSCTASLQVSDRTVPPSFTRKLKETNGLSGSVVYMECKVY 4371

Qy 970 GIPVPKYVWFKDGKQIISKRNHCKMRREGDGTCSLHIESTTSDDDGNYTTMAANPOGRIS 1029
Db 4372 GSPPIVSFWFHEGNEISGRKY--QTTLDNTCALTVMLEESDSDGYTCIATNMAGSDE 4429

Qy 1030 CSGHLMVQSLPIRSLTSAGOSHGRSRVORDEKEPFRPHFLQAPGDMVAHGEGL 1089
Db 4430 CSAPLTV-----REP-----PSFVQKQPDMDVLGTN 4456

Qy 1090 CRLDCKVSGLPPPELTWLLNGQVLP--DASHKMLVRETGVHSLLDPLTQRDAGTYKCIA 1148
Db 4457 VTFTSIVKGTGPPFSVSWFKGSSSELVPGDRCNVSL--EDSVAEELFVDVTSQSGEYTCIV 4514

Qy 1149 TNKTQNSFSLSLSVAKREVKKAPV--ILEKLNQGVPEGHVPRLCVRVIGMPPVPVYWK 1207
Db 4515 SNEAGKASCCTHLYI-----KAPAKFVKRLNDYSIEKGKPLILEGTTFTPTPPISVTWKK 4568

Qy 1208 DNETI-PCRTERISMHDQDTGYACLIQIAPK--SDAGWYTLISAKNEAGIVSCVTCARLDIYA 1265
Db 4569 NGINVTPSQRGNIT---TTEKSPILEIPSTVEDAGQYCNICYENASGKDCSAQILI-- 4622

Qy 1266 QWHHOIPIPP-----MSVRPSGSRYSGLTSK---GLDIFSASFSSMESTWYVSSSR 1312
Db 4623 -----LEPPYFVVKOLEPVKVSVDGSASLQCOLAGTPEIGVSWYKGDTKLRPTTTYKMHFR 4677
```


Db 1672 STLTRSPSVENIAKYQAAQAPETHRATTGTSQFVANIAQVHNPSQPERQVFTKETPKID 1731
QY 803 -----SG----- 804
Db 1732 RIDDILATPHEKVKVVKOTEVITSDVKNYQEQQKQOQVTSVTVTTSGEQWVOHQH 1791
QY 805 -----NQFQPRCVSP-----IPVSPTSRI 823
Db 1792 QDFVEPHKSTITVKKLDIEQQNINOQFAQPEQKQHITTEKIEBDLHRESKIPVKREPQT 1851
QY 824 QNPVAFSS-----VLP-----LPAIPTNAMXLP-----RSAPSMSQGLAKNT 865
Db 1852 TTTVTETTSGEQWQNSHODFQPHLSTTTVKKLDIDEGHLLNYQTAELQOOQ-VKTRT 1910
QY 866 KSPQPVNDNIR-ETKANVIRDLGKKITFSDV-----RP----- 898
Db 1911 KEE---NESIRNESKIPVAREPKTTTTVTTSGEQWQVNHQDLRRPQSTVTVKRLD 1967
QY 899 -----NQOEYKISFE-----QRLMNEIEFRLETP----- 924
Db 1968 IEGGEHVHDTQLLRNQOQVVTKTEENDQRMKNESRIPVRRPEKTTTTVTETTSGEW 2027
QY 925 -----VDE-----SDDE--IQH----- 934
Db 2028 VQNSHMDAANPQKSTVTVRRLGIDEETTKTSARPISSEIRROQDEESSVRHRDSFIQAS 2087
QY 935 -----DEIPT-----GKCIAPIDKRL-KHFRVTEGS 960
Db 2088 DVEGFTWDGAYTSAPTPPPQPIHRSTAENDMQRIGLSRTTTEPEFKAFERYTVBEGG 2147
QY 961 PVFTCKIVGIPVKVYFKDGKQISKRNECHKMRREGDGTCSLHIESTTSDDDGNTYIM 1020
Db 2148 RIAIECILVGNPKPAARFFFNKNQVTEKSEBFLKICHVND-TYSIIISPAKLEHAGYKMI 2206
QY 1021 AANPOGRISCSGHLWOSLP---IRSR-LTSAGQSHGRGRVQER----- 1061
Db 2207 AENKRG-----VTESLVLHVRRSLQYQOKKQOETRMQOQARNAGTTEYTTVEE 2258
QY 1062 -----DKPQLQBRFR-----PHFLQAP 1079
Db 2259 FMPEYEQRLPKHEASKLSTPPPAKRIELEHRRKDEHELYDLEGKKKGCHPHFTQTL 2318
QY 1080 GDMVAHEGRCLRDCKYSGLPPPELTWLLNGQVLPDASHKMLVRETGVHSLLDPLTQR 1139
Db 2319 VSTVVAQGEATFEGIVTGVPAPNVWENDGRPL-DLKDIRVSNIGRVSLSNFQNCQLS 2376
QY 1140 DAGTYKCIATNKTCQNSFSELSLVVAKEVKAPVILEKLQNCQVPEGHPVRLCERVIGMP 1199
Db 2377 HVGYMCTAKNDGVSATSSAQLVVRPKTI--APDFIQRLISEEIEEGSQLKWTVTVTGDP 2434
QY 1200 PVFVYWKDNETIP-CTRERISMHODTGYACLLIQPAKKSADAGWYTLISAKNEAGIVSCT 1258
Db 2435 MPKVIWNRDGEIIPDCREVRIVDGD--GYHSLVIVKVEGADSGQFTCLAENIAGEARST 2492
QY 1259 ARLDIYAQWHHQIIPPMPSVRPSGRYS 1286
Db 2493 ADL-----VVRPQGTAPGN 2506

RESULT 10
T20992

hypothetical protein F15G9.4a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T20992; T24733

R:Sulston, J.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19355

A:Accession: T20992

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <W1>

A:Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone F15G9

R:Kershaw, J.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19929

A:Accession: T24733

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <W12>

A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone T09B9

C:Genetics:

A:Gene: CESP:F15G9.4a

A:Map position: X

A:Introns: 85/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184

A:Exons: 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 303

1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 50

Query Match

Best Local Similarity 6.8%; Score 468.5; DB 2; Length 5175;

Matches 248; Conservative 161; Mismatches 448; Indels 299; Gaps 48;

QY 268 PPR--FTOKLRSREVPECTRVOLDIVGVIPPPQVRYWYCEK--ELENSPDIHIVQAGNL 323

Db 2286 PPTIIMLDKDKNKTAVEHSTVTLSCPATGKPEPDITWFKDGEATHENIAD--IIPNGEL 2343

QY 324 --HSLTIAAEFEEDTGRYSCFASNIYTDSTAETIYEGVSSDSEGDPNKEENNRIOQP 381

Db 2344 NGNQLKTRIKEGDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDGIPISEYQ--QNE 2401

QY 382 NEVSSPPTTSVAPPA---VPQAHLVAQPRVATIOOCQS-----PT 420

Db 2402 RVVISCPIYAR--PPAKITWLKAGKQLQSDKFVKTSAANGKLYLFKLRETDSSKYTCIAT 2459

QY 421 NYLOGLDCK---PLIAAPVETK--MLQNLASSEGOLVVVFECRVKGPAPVAYRECTL 474

Db 2460 NE-AGTDKRDPKVSMVAPSEDEPNIVRRITVNSGNPSTLHCPKAGSPSPITWLKDNA 2518

QY 475 IEDSPDFRILQKKPRMAEPETICTLVIAEVEAFDSDGCTCTASNKYGTVSIAQLHVRG 534

Db 2519 IEPNDRVYFFDAGRQ-----LOISKTEGSDOGRYTCIATNSV-----G 2556

QY 535 NEDLSN-----NGSLHSANSTNLAAIEPQPPSPHSEPPSPVEOPPKKLEGVLVN 584

Db 2557 SDDLENTLEVIPIPPVIDGEREA---VAVIEGFSELSFCDSDNSTGVDVEMQKDLGTIN 2611

QY 585 HNEPRSSRIGRLRVHFLPEDDKG---SEASSEAGVVTTRQTRPDSQOERFNGQATKP 640

Db 2612 QDTLRGDSFI-----QIPSGKKMSFLSARKSDSGRYTCIVRNP-----AGEARKLF 2658

QY 641 EPSFPVKEPPVLAAPKPLDSTQLQOL-----HNQVLLLEQHLQNPSPSS 684

Db 2659 D--FAVNDPPSI--SDELSANIQTIVPYYPVEINCVVSGSPHPKV---YWLFDKPLE 2710

QY 685 PKPEFFXMT-----VLNSNAPPAVT--TSKQVKAPSSQTFSLAR---PKYFFSTWTT 733

Db 2711 PDSAAVELTNNGETILKIVRSQVEHAGTYTCEAQNNGVARKARKDLVVRTAPPHEKEREVE 2770

QY 734 AATV-----APSSSPVFTL--SSTPQTIQRTV-----SKESLLVSHPSVQTKSP 775

Db 2771 VARVGDITMLTCAESSVPLSSVYVWHADESQVNGSVITSKYAANEKTLNVT--NIQLDDE 2828

QY 776 G-----GLS-----IQNEPLP-----GTEPTTPTPTESI 801

Db 2829 GFYCTAVNEAGITKKFKFLIVETPYPLDQOKLYPIILGKRLDLCDSATGTPPTIIFM 2888

QY 802 PSQNFQPRCVSPIPVSPTSRIQNP---VAFLSVLPSPAIPTNAMXLPSPAPSPSP 858

Db 2889 KDGKRLNESDEVDI--IGSTLVLDNPNQKEVEGRYTCIAENKAGRSEKDMVVLPPKLSK 2947

QY 859 GLAKNNTKSPQPVNDNDNRETKNVIRDL-GKKITFSDVRNQOYKIKISSFEQRLMNEIE 917

Db 2948 EWINVEVOAGDPL-----TLECPIDTSGVHTWS-----RQFGKQGLDMRAQSSSD 2995

Qy 918 ----FRLETPVDESDDDEIQHDEIPTGKCIA-----PIFKRLKHF 954
Db 2996 KSKLYIMQATPEDAD-----SYSCIAVNDAGGAEAVFQVTVNTPPKIFGDSFSTT 3045
Qy 955 RVTEGSPVTFCKIVGIPVVKVYFKDGKQISKNEHCKMRREG-----DGTCSLHIESTT 1010
Db 3046 EIVADTTLEIPCRTEGIPPPISWFLDGKPI-----LEMPGVYTKQGSLSLRIDNIK 3097
Qy 1011 SDDGNTYMAANPQGRISCSGHLWVQSLPIRSRLTSAGQSHRGRSRVQERDKPELQERF 1070
Db 3098 PNQGRYTCVAENKAGRAEQDTYVEI-----SEP----- 3126
Qy 1071 FRPFLQAPGDMAHEGRCLRDLCKVSGLPPELTWLLNGQVLPDPASHKMLVRETGVHS 1130
Db 3127 --PRVMASEVMRVVEGRQTTIRCEVEGPNPEVNVNWKLDGEPYTS-----LLQFSTKLSY 3180
Qy 1131 LLIDPLTQRDAGTYKCIATNKTGNSFSLSVVAKKAPVILEKLQNGCVPEGHVPR 1190
Db 3181 LHLRETTLADGGTYTCIATNKAGESQTTDV-----EVLVPPRIEDEERVVLQGEKNTYM 3235
Qy 1191 LECRVIGMPPPVFVYKKDNETIPTCTRERISMHQDTTGYACLLIOPAKKSDAGWTLTSAKN 1250
Db 3236 VHCQVTRGPVYVTKRNGKEI-----EQFN-----PVLHIRNATRADEGKYSCIASN 3283
Qy 1251 EAGVSCITARDIYQAQ 1266
Db 3284 EAGTAVADFLIDVETK 3299
RESULT 11
T43290
hemocentin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43290; T20993; T24734
R:Vogel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A:Description: Hemocentin is required for hemidesmosome mediated cell adhesion and germ-
A:Reference number: 222396
A:Accession: T43290
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <VQ>
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20993
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1
; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1
Query Match 6.8%; Score 468.5; DB 2; Length 5198;
Best Local Similarity 21.58; Pred. No. 1.9e-14;
Matches 248; Conservative 161; Mismatches 448; Indels 299; Gaps 48;

Qy 268 PPR--FTQKLSRREVECTRVQLDCIVVGIPIPPQVRWVCEGK--ELENSPDHIVQAGNL 323
Db 2286 PPTIIMLDKDKNKTAVEHSTVTLSCPATGKPEPDITWFKDGEATHENIAD--IIPNGEL 2343
Qy 324 --HSLTTAAFEEDTGRYSCFASNIYGTDSYSAEIIYEGVSSSDSEDPNKEENNRITQKP 381
Db 2344 NGNOLKITRIKREGAGKYTCEDANSAGSVEQDVNVNVTIPKIEKDGIPSDYESQ--QNE 2401
Qy 382 NEVSSPPTTSVAVIPPA---VPOAHLVAOPRVATIOQCQS-----PT 420
Db 2402 RVVJISCPYAR--PPAKITWLKAGKPLQSDKFVKTSANGQKLYLFKURETOSSKYTCTIAT 2459
Qy 421 NYLOGLDGK----PIIAAPVETK--MLQNLASSEGQLVVFEKRVKGPAPSPKVEVREGTL 474
Db 2460 NE-AGTKRDPKVKMLVAPSPDEPNIVRRITVNSGNPSTLHCPAKGSPSPITWLKDGNA 2518
Qy 475 IEDSPDFRILQKPRMAPEEICTLVIAEVAEDSGCTCTTASNKYGTGTVSSIAQLHVHG 534
Db 2519 IEPNDRVVFFDAGRQ-----LQISKTEGSDQGRYTCIATNSV-----G 2556
Qy 535 NEDLSN-----NGSLHSANSTTNLAAIEPQSPPHSEPPSVEQPPKPLEGLVLVN 584
Db 2557 SDDLENTLEIIPVIDGERREA-----VAVIEGFSSELCFCDNSSTGVVDVEMQXDLGIN 2611
Qy 585 HNEPRSSRIGLRVHFNLPEDDKG-----SEASSEAGVVVTRQTRPDSAXQERFNGQATKTP 640
Db 2612 QDTLRGDSFI-----QIPSSGKKMSLSARKSDSGRYTCIVRNP-----AGEARKLF 2658
Qy 641 EPSFPVKPEPPVLAKPKLDSTQLQOL-----HNOVLLEQHQLOLQNPSPSS 684
Db 2659 D--FAVNDPPSI--SDELSANIQTIVPYVPEINCVVVSGSPHPKV---YVLFDDKPLE 2710
Qy 685 PKRFPFXMT-----VLNSNAPPVAVT-TSKQVKAPASSQTFESLAR---PKYFPSTNTT 733
Db 2711 PDSAAYELTNNGETLKVRSQVEHAGTYTCEAQNNGKARKDFLVRVTAPHPHKEKEEV 2770
Qy 734 AATV-----APSSSPVFTL--SSTPQTQRTV-----SKESLLVSHPSVQTKSP 775
Db 2771 VARVGDMILLTCNAESSVPLSSVYVHWAHDESVQNGVITSKYAANEKTLNVT--NIQLDDE 2828
Qy 776 G-----GLS-----IQNEPLPP-----GPTETPPPTFFSI 801
Db 2829 GFYCTAVNEAGITKFKFLIVETPYFLDQOKLYPIILGKRLTLDGCSATGTPPTTILFM 2888
Qy 802 PSGNQFQPRCVSPIPVSPTSRIQNP--VAFLSVLSPLPAIPPTNAMXLPSPASPMPSQ 858
Db 2889 KDGRLNESDEVDI-IGSTLVIDNPQKEVEGRTYCIENKAGRSEKDMVAVLLPKLSK 2947
Qy 859 GLAKKNTKSPQPVNDNIRETKNAVIRDL-GKKTTFSDVRPNQOEYKISSFEQRLMNEIE 917
Db 2948 EWINVEVQAGDPL-----TLECPIDETSGVHITWS-----RQFGDKGOLDMRAQSSSD 2995
Qy 918 ----FRLETPVDESDDDEIQHDEIPTGKCIA-----PIFKRLKHF 954
Db 2996 KSKLYIMQATPEDAD-----SYSCIAVNDAGGAEAVFQVTVNTPPKIFGDSFSTT 3045
Qy 955 RVTEGSPVTFCKIVGIPVVKVYFKDGKQISKNEHCKMRREG-----DGTCSLHIESTT 1010
Db 3046 EIVADTTLEIPCRTEGIPPPISWFLDGKPI-----LEMPGVYTKQGSLSLRIDNIK 3097
Qy 1011 SDDGNTYMAANPQGRISCSGHLWVQSLPIRSRLTSAGQSHRGRSRVQERDKPELQERF 1070
Db 3098 PNQGRYTCVAENKAGRAEQDTYVEI-----SEP----- 3126
Qy 1071 FRPFLQAPGDMAHEGRCLRDLCKVSGLPPELTWLLNGQVLPDPASHKMLVRETGVHS 1130
Db 3127 --PRVMASEVMRVVEGRQTTIRCEVEGPNPEVNVNWKLDGEPYTS-----LLQFSTKLSY 3180
Qy 1131 LLIDPLTQRDAGTYKCIATNKTGNSFSLSVVAKKAPVILEKLQNGCVPEGHVPR 1190
Db 3181 LHLRETTLADGGTYTCIATNKAGESQTTDV-----EVLVPPRIEDEERVVLQGEKNTYM 3235
Qy 1191 LECRVIGMPPPVFVYKKDNETIPTCTRERISMHQDTTGYACLLIOPAKKSDAGWTLTSAKN 1250

Db 3236 VHCQVTRPVYVWKRNGKEI-----EQFN-----PVLHIRNATRADEGKYSCIASN 3283
QY 1251 EAGIVSCTARLDIYAO 1266
Db 3284 EAGTAVADFLIDVETK 3299
RESULT 12
T42634
connectin/titin - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
R;Accession: T42634
R;Yaajima, H.; Ohtsuka, H.; Kume, H.; Endo, T.; Kimura, S.; Maruyama, K.
Zool. Sci. 13, 119-123, 1996
A>Title: Molecular cloning of a partial cDNA clone encoding the C terminal region of chi
A:Reference number: 222220; MUID:96261234; PMID:8688806
A:Accession: T42634
A:Molecule type: mRNA
A:Residues: 1-1021 <VAV>
A:Cross-references: EMBL:D83008; NID:g1181556; PIDN:BAA11678.1; PID:g1777308
A:Experimental source: breast muscle; clone CC4
C:Keywords: skeletal muscle

Query Match 6.2%; Score 429.5; DB 2; Length 1021;
Best Local Similarity 18.7%; Pred. No. 2e-13;
Matches 231; Conservative 138; Mismatches 379; Indels 487; Gaps 37;

QY 170 SKRIRPRACKNHKSLKSONKVMQESSFSDLSERRERSVPPIPAD----- 218
Db 58 SELIRPRSRPPEESERSP-----TPERTRPRSPSPVSTERLSRFERMAR 103
QY 219 -----TRNEVNHAEQEAQKREAEQAASEAGDTPGSPFSLYIEEP--LQPPRFT 272
Db 104 FDIERSYEMKSAKTKQTKMERKE-----VLTQQPPTLDHAPRIT 144
QY 273 QKLSREVPECTRVQDCIVVIGIPPQVRWYCEGKELENSPDHIVQAGNLSHTIAEAF 332
Db 145 LRMSHRVPCGHNTFRILNVQSKTADVKYTHNGIELQESSKIHFNTSGVLTLELDCH 204
QY 333 EEDTGRYSCFASNIYGTSTSAEIIYIEGVSSSDSEGDPNKEEMNRIOKN-----EVS 385
Db 205 IDDSGTAVCTNKGECSDVATLDVTGGDYTYTSSQRDEEVPRSLPDLTRTEAVYS 264
QY 386 S-PPTTSVAVPPVPAQHLVAQPR--VATIQQCQSPNYLQGLDGKPI----- 431
Db 265 SFKKATAAEASSSVREYKSEVSATRESLLSYEHASSEEKITASEEKSLSEERTVHKAFKS 324
QY 432 -IAAPVFTKMLQNLASSEGOLVVEECRVKGAPSPKVEYREGTLIEDSPDFRILQKKPRS 490
Db 325 TLPATILTKP-RSITVSEGETARFSCDVGEPATITWVRAGQIVSRREFQIRTVQKS 383
QY 491 MAEPEICTLVIAEVFAEDSCFTCTASNKYGTVSSIAQLHVRGNEDLSNGLHSANST 550
Db 384 -----TFEISLVQIADEGSYT----- 399
QY 551 TNLAALIEQPPSPHSEPPSVEQPPKPLEGVLVNHNENRSPSSRIGLRVHFNLPEDDKGSE 610
Db 400 -----YVVENSEGR----- 408
QY 611 ASSEAGVVTTROTSPDSXQERFNGQATKTPSPSPVKEPPVPLAKPKLDSTQLQOLHNQV 670
Db 409 -----QEAH--- 412
QY 671 LLEQHLQNPPSPKPEFFXMTVLNSNAPPVATTSKXQKAPSSQTFSLARPKYFFEST 730
Db 413 -----FTLVQRKRIPEKAITSPRIKSP----- 436
QY 731 NTTAATVAPSSSPVTLSSPTQTQRTVSKESLLVSHPSVQTKSPGGGLSIONEPLPGPT 790
Db 437 ----- 436

QY 791 EPTPPPTFSIPSGNOFQRCVSPIDPSPTRSQIONPVAFSLSSVLPSPAIPTTNAMKLPR 850
Db 437 -----EPRVKSPPEVKSPE-----PI 458
QY 851 SAPSMPSOGLAKNKTQPPVNDNIRETKNAVIRDLGKKTTFSDVRPNQOYKISSPEQ 910
Db 459 STPS-----KAKSPGD----- 470
QY 911 RLMNEIERLERLTPVDESDDEIOHDEIPTGKCIAPIDKRLKHLFRVTGSPVTFCTKIVG 970
Db 471 -----KTAPVEKV-----OLPTAS--PPKIKEQLK--AETLGDKVKLSCAVES 509
QY 971 --IPVPKVYFKDCQKQISKRNEHCKMRREGDGTCSLHTESTTDDDGNYT-----IM 1020
Db 510 SVLSIREVAVWYKDGKCL-KEDHHEKFHYAAGTYELAKHNLTESDKGEYTCIEIMEGGIS 568
QY 1021 AANPQ--GRISCSGHLMVQSLPIRSRLTSAGO-----SHRGRSRVQERDKPELQE----- 1068
Db 569 KTNFQFTQGVFKNIHSQVSVSETPKSVKGDGKVLAVSTQKSSAATEEKAIEIEVIKKS 628
QY 1069 -----REFRPHFLQAPCDMVAHREGRLCRDCKVSGLPPPELTWLLNGQVLPDASHKML 1122
Db 629 IVTEVDVQLOAEIRASSTQMTVSEGVTKLANIPG--ASEVKVWLNGMELRNSDDYRYG 686
QY 1123 VRETVGSHLLDPLTQORDAGTYKCIATNKTGNSFSLE--LSWAKEVKKAPVILEKLQNC 1181
Db 687 ISGSN-HLTITKKASNKDEGILTC--EKGTEGTIKCOYVLTFSKEPSNEFAITQPKSQ 743
QY 1182 GVPEGHPVRLCEVRIGMPPPVYVWKKDNETICTRERISMHDQTT---GYACLLIQPAKK 1238
Db 744 NVNEGQDVLFTCEVSGDPSPEVWLNNQPI-----AVSSHMRATRSTKNTYSLEIRNAV 798
QY 1239 SDAGWYTLISAKNEAGIVSCTARLDIYAQWHQIIPPMVS--RPSG--SPYGLSLTSKGLDI 1294
Db 799 SDTKYTVKAKNYHGQCSATASLTVPFLIEE---PPKEVVLKTSGDASHMESFSQSQFM 855
QY 1295 -----FSAF---SSMEBTWVYSCSSRSVVESE 1319
Db 856 AASKQEAASFSSFSSTSTETKFASSAKSMSSMKE 890

RESULT 13
A88852
protein unc-22 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
R;Accession: A88852
R;anonymus, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: A88852
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN00022
C:Genetics:
A:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homolog

Query Match 5.8%; Score 402; DB 2; Length 6831;
Best Local Similarity 18.3%; Pred. No. 4.7e-11;
Matches 288; Conservative 153; Mismatches 532; Indels 600; Gaps 50;
QY 204 ERRERSSVIPADTRDNEVNHAEQ--QEAQRRAEQAAAEAG-GDTPGSSPSSLYY 261
Db 5141 EIREYGSGLTWVADSYNVREPEFTVKLREFNDYEFVRVAVINAAKGIPSLSPSPGIKI-- 5198
QY 262 EEPGQPPRFTQKLSREVPECTRVQDCIVVIGIPPQVRWYCEGKELENSPDHIVQAG 321

A; Experimental source: var. Bristol
R; White, S.
submitted to the EMBL Data Library, May 1996
A; Reference number: Z20442
A; Accession: Z27934
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 'MGIPGKKCKO', 19-6839 <WIL>
A; Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:ZK617.1a
A; Experimental source: clone ZK617
R; Harris, B.
submitted to the EMBL Data Library, May 1996
A; Reference number: Z20458
A; Accession: Z28030
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 'MGIPGKKCKO', 19-6839 <W2>
A; Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a
A; Experimental source: clone ZK829
C; Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
C; Genetics:
A; Gene: unc-22; CESP:ZK617.1a
A; Map position: 4
A; Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 601/2; 669/1; 677/1; 6808/3
C; Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;
C; Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/threonine kinase
F: 806-898, 899-990, 991-1083, 1084-1175, 1178-1273, 1474-1567, 1770-1864, 2066-2158, 2358-2450, 2451-2510, 2511-2512, 2513-2514, 2515-2516, 2517-2518, 2519-2520, 2521-2522, 2523-2524, 2525-2526, 2527-2528, 2529-2530, 2531-2532, 2533-2534, 2535-2536, 2537-2538, 2539-2540, 2541-2542, 2543-2544, 2545-2546, 2547-2548, 2549-2550, 2551-2552, 2553-2554, 2555-2556, 2557-2558, 2559-2560, 2561-2562, 2563-2564, 2565-2566, 2567-2568, 2569-2570, 2571-2572, 2573-2574, 2575-2576, 2577-2578, 2579-2580, 2581-2582, 2583-2584, 2585-2586, 2587-2588, 2589-2590, 2591-2592, 2593-2594, 2595-2596, 2597-2598, 2599-2600, 2601-2602, 2603-2604, 2605-2606, 2607-2608, 2609-2610, 2611-2612, 2613-2614, 2615-2616, 2617-2618, 2619-2620, 2621-2622, 2623-2624, 2625-2626, 2627-2628, 2629-2630, 2631-2632, 2633-2634, 2635-2636, 2637-2638, 2639-2640, 2641-2642, 2643-2644, 2645-2646, 2647-2648, 2649-2650, 2651-2652, 2653-2654, 2655-2656, 2657-2658, 2659-2660, 2661-2662, 2663-2664, 2665-2666, 2667-2668, 2669-2670, 2671-2672, 2673-2674, 2675-2676, 2677-2678, 2679-2680, 2681-2682, 2683-2684, 2685-2686, 2687-2688, 2689-2690, 2691-2692, 2693-2694, 2695-2696, 2697-2698, 2699-2700, 2701-2702, 2703-2704, 2705-2706, 2707-2708, 2709-2710, 2711-2712, 2713-2714, 2715-2716, 2717-2718, 2719-2720, 2721-2722, 2723-2724, 2725-2726, 2727-2728, 2729-2730, 2731-2732, 2733-2734, 2735-2736, 2737-2738, 2739-2740, 2741-2742, 2743-2744, 2745-2746, 2747-2748, 2749-2750, 2751-2752, 2753-2754, 2755-2756, 2757-2758, 2759-2760, 2761-2762, 2763-2764, 2765-2766, 2767-2768, 2769-2770, 2771-2772, 2773-2774, 2775-2776, 2777-2778, 2779-2780, 2781-2782, 2783-2784, 2785-2786, 2787-2788, 2789-2790, 2791-2792, 2793-2794, 2795-2796, 2797-2798, 2799-2800, 2801-2802, 2803-2804, 2805-2806, 2807-2808, 2809-2810, 2811-2812, 2813-2814, 2815-2816, 2817-2818, 2819-2820, 2821-2822, 2823-2824, 2825-2826, 2827-2828, 2829-2830, 2831-2832, 2833-2834, 2835-2836, 2837-2838, 2839-2840, 2841-2842, 2843-2844, 2845-2846, 2847-2848, 2849-2850, 2851-2852, 2853-2854, 2855-2856, 2857-2858, 2859-2860, 2861-2862, 2863-2864, 2865-2866, 2867-2868, 2869-2870, 2871-2872, 2873-2874, 2875-2876, 2877-2878, 2879-2880, 2881-2882, 2883-2884, 2885-2886, 2887-2888, 2889-2890, 2891-2892, 2893-2894, 2895-2896, 2897-2898, 2899-2900, 2901-2902, 2903-2904, 2905-2906, 2907-2908, 2909-2910, 2911-2912, 2913-2914, 2915-2916, 2917-2918, 2919-2920, 2921-2922, 2923-2924, 2925-2926, 2927-2928, 2929-2930, 2931-2932, 2933-2934, 2935-2936, 2937-2938, 2939-2940, 2941-2942, 2943-2944, 2945-2946, 2947-2948, 2949-2950, 2951-2952, 2953-2954, 2955-2956, 2957-2958, 2959-2960, 2961-2962, 2963-2964, 2965-2966, 2967-2968, 2969-2970, 2971-2972, 2973-2974, 2975-2976, 2977-2978, 2979-2980, 2981-2982, 2983-2984, 2985-2986, 2987-2988, 2989-2990, 2991-2992, 2993-2994, 2995-2996, 2997-2998, 2999-3000, 3001-3002, 3003-3004, 3005-3006, 3007-3008, 3009-3010, 3011-3012, 3013-3014, 3015-3016, 3017-3018, 3019-3020, 3021-3022, 3023-3024, 3025-3026, 3027-3028, 3029-3030, 3031-3032, 3033-3034, 3035-3036, 3037-3038, 3039-3040, 3041-3042, 3043-3044, 3045-3046, 3047-3048, 3049-3050, 3051-3052, 3053-3054, 3055-3056, 3057-3058, 3059-3060, 3061-3062, 3063-3064, 3065-3066, 3067-3068, 3069-3070, 3071-3072, 3073-3074, 3075-3076, 3077-3078, 3079-3080, 3081-3082, 3083-3084, 3085-3086, 3087-3088, 3089-3090, 3091-3092, 3093-3094, 3095-3096, 3097-3098, 3099-3100, 3101-3102, 3103-3104, 3105-3106, 3107-3108, 3109-3110, 3111-3112, 3113-3114, 3115-3116, 3117-3118, 3119-3120, 3121-3122, 3123-3124, 3125-3126, 3127-3128, 3129-3130, 3131-3132, 3133-3134, 3135-3136, 3137-3138, 3139-3140, 3141-3142, 3143-3144, 3145-3146, 3147-3148, 3149-3150, 3151-3152, 3153-3154, 3155-3156, 3157-3158, 3159-3160, 3161-3162, 3163-3164, 3165-3166, 3167-3168, 3169-3170, 3171-3172, 3173-3174, 3175-3176, 3177-3178, 3179-3180, 3181-3182, 3183-3184, 3185-3186, 3187-3188, 3189-3190, 3191-3192, 3193-3194, 3195-3196, 3197-3198, 3199-3200, 3201-3202, 3203-3204, 3205-3206, 3207-3208, 3209-3210, 3211-3212, 3213-3214, 3215-3216, 3217-3218, 3219-3220, 3221-3222, 3223-3224, 3225-3226, 3227-3228, 3229-3230, 3231-3232, 3233-3234, 3235-3236, 3237-3238, 3239-3240, 3241-3242, 3243-3244, 3245-3246, 3247-3248, 3249-3250, 3251-3252, 3253-3254, 3255-3256, 3257-3258, 3259-3260, 3261-3262, 3263-3264, 3265-3266, 3267-3268, 3269-3270, 3271-3272, 3273-3274, 3275-3276, 3277-3278, 3279-3280, 3281-3282, 3283-3284, 3285-3286, 3287-3288, 3289-3290, 3291-3292, 3293-3294, 3295-3296, 3297-3298, 3299-3300, 3301-3302, 3303-3304, 3305-3306, 3307-3308, 3309-3310, 3311-3312, 3313-3314, 3315-3316, 3317-3318, 3319-3320, 3321-3322, 3323-3324, 3325-3326, 3327-3328, 3329-3330, 3331-3332, 3333-3334, 3335-3336, 3337-3338, 3339-3340, 3341-3342, 3343-3344, 3345-3346, 3347-3348, 3349-3350, 3351-3352, 3353-3354, 3355-3356, 3357-3358, 3359-3360, 3361-3362, 3363-3364, 3365-3366, 3367-3368, 3369-3370, 3371-3372, 3373-3374, 3375-3376, 3377-3378, 3379-3380, 3381-3382, 3383-3384, 3385-3386, 3387-3388, 3389-3390, 3391-3392, 3393-3394, 3395-3396, 3397-3398, 3399-3400, 3401-3402, 3403-3404, 3405-3406, 3407-3408, 3409-3410, 3411-3412, 3413-3414, 3415-3416, 3417-3418, 3419-3420, 3421-3422, 3423-3424, 3425-3426, 3427-3428, 3429-3430, 3431-3432, 3433-3434, 3435-3436, 3437-3438, 3439-3440, 3441-3442, 3443-3444, 3445-3446, 3447-3448, 3449-3450, 3451-3452, 3453-3454, 3455-3456, 3457-3458, 3459-3460, 3461-3462, 3463-3464, 3465-3466, 3467-3468, 3469-3470, 3471-3472, 3473-3474, 3475-3476, 3477-3478, 3479-3480, 3481-3482, 3483-3484, 3485-3486, 3487-3488, 3489-3490, 3491-3492, 3493-3494, 3495-3496, 3497-3498, 3499-3500, 3501-3502, 3503-3504, 3505-3506, 3507-3508, 3509-3510, 3511-3512, 3513-3514, 3515-3516, 3517-3518, 3519-3520, 3521-3522, 3523-3524, 3525-3526, 3527-3528, 3529-3530, 3531-3532, 3533-3534, 3535-3536, 3537-3538, 3539-3540, 3541-3542, 3543-3544, 3545-3546, 3547-3548, 3549-3550, 3551-3552, 3553-3554, 3555-3556, 3557-3558, 3559-3560, 3561-3562, 3563-3564, 3565-3566, 3567-3568, 3569-3570, 3571-3572, 3573-3574, 3575-3576, 3577-3578, 3579-3580, 3581-3582, 3583-3584, 3585-3586, 3587-3588, 3589-3590, 3591-3592, 3593-3594, 3595-3596, 3597-3598, 3599-3600, 3601-3602, 3603-3604, 3605-3606, 3607-3608, 3609-3610, 3611-3612, 3613-3614, 3615-3616, 3617-3618, 3619-3620, 3621-3622, 3623-3624, 3625-3626, 3627-3628, 3629-3630, 3631-3632, 3633-3634, 3635-3636, 3637-3638, 3639-3640, 3641-3642, 3643-3644, 3645-3646, 3647-3648, 3649-3650, 3651-3652, 3653-3654, 3655-3656, 3657-3658, 3659-3660, 3661-3662, 3663-3664, 3665-3666, 3667-3668, 3669-3670, 3671-3672, 3673-3674, 3675-3676, 3677-3678, 3679-3680, 3681-3682, 3683-3684, 3685-3686, 3687-3688, 3689-3690, 3691-3692, 3693-3694, 3695-3696, 3697-3698, 3699-3700, 3701-3702, 3703-3704, 3705-3706, 3707-3708, 3709-3710, 3711-3712, 3713-3714, 3715-3716, 3717-3718, 3719-3720, 3721-3722, 3723-3724, 3725-3726, 3727-3728, 3729-3730, 3731-3732, 3733-3734, 3735-3736, 3737-3738, 3739-3740, 3741-3742, 3743-3744, 3745-3746, 3747-3748, 3749-3750, 3751-3752, 3753-3754, 3755-3756, 3757-3758, 3759-3760, 3761-3762, 3763-3764, 3765-3766, 3767-3768, 3769-3770, 3771-3772, 3773-3774, 3775-3776, 3777-3778, 3779-3780, 3781-3782, 3783-3784, 3785-3786, 3787-3788, 3789-3790, 3791-3792, 3793-3794, 3795-3796, 3797-3798, 3799-3800, 3801-3802, 3803-3804, 3805-3806, 3807-3808, 3809-3810, 3811-3812, 3813-3814, 3815-3816, 3817-3818, 3819-3820, 3821-3822, 3823-3824, 3825-3826, 3827-3828, 3829-3830, 3831-3832, 3833-3834, 3835-3836, 3837-3838, 3839-3840, 3841-3842, 3843-3844, 3845-3846, 3847-3848, 3849-3850, 3851-3852, 3853-3854, 3855-3856, 3857-3858, 3859-3860, 3861-3862, 3863-3864, 3865-3866, 3867-3868, 3869-3870, 3871-3872, 3873-3874, 3875-3876, 3877-3878, 3879-3880, 3881-3882, 3883-3884, 3885-3886, 3887-3888, 3889-3890, 3891-3892, 3893-3894, 3895-3896, 3897-3898, 3899-3900, 3901-3902, 3903-3904, 3905-3906, 3907-3908, 3909-3910, 3911-3912, 3913-3914, 3915-3916, 3917-3918, 3919-3920, 3921-3922, 3923-3924, 3925-3926, 3927-3928, 3929-3930, 3931-3932, 3933-3934, 3935-3936, 3937-3938, 3939-3940, 3941-3942, 3943-3944, 3945-3946, 3947-3948, 3949-3950, 3951-3952, 3953-3954, 3955-3956, 3957-3958, 3959-3960, 3961-3962, 3963-3964, 3965-3966, 3967-3968, 3969-3970, 3971-3972, 3973-3974, 3975-3976, 3977-3978, 3979-3980, 3981-3982, 3983-3984, 3985-3986, 3987-3988, 3989-3990, 3991-3992, 3993-3994, 3995-3996, 3997-3998, 3999-4000, 4001-4002, 4003-4004, 4005-4006, 4007-4008, 4009-4010, 4011-4012, 4013-4014, 4015-4016, 4017-4018, 4019-4020, 4021-4022, 4023-4024, 4025-4026, 4027-4028, 4029-4030, 4031-4032, 4033-4034, 4035-4036, 4037-4038, 4039-4040, 4041-4042, 4043-4044, 4045-4046, 4047-4048, 4049-4050, 4051-4052, 4053-4054, 4055-4056, 4057-4058, 4059-4060, 4061-4062, 4063-4064, 4065-4066, 4067-4068, 4069-4070, 4071-4072, 4073-4074, 4075-4076, 4077-4078, 4079-4080, 4081-4082, 4083-4084, 4085-4086, 4087-4088, 4089-4090, 4091-4092, 4093-4094, 4095-4096, 4097-4098, 4099-4100, 4101-4102, 4103-4104, 4105-4106, 4107-4108, 4109-4110, 4111-4112, 4113-4114, 4115-4116, 4117-4118, 4119-4120, 4121-4122, 4123-4124, 4125-4126, 4127-4128, 4129-4130, 4131-4132, 4133-4134, 4135-4136, 4137-4138, 4139-4140, 4141-4142, 4143-4144, 4145-4146, 4147-4148, 4149-4150, 4151-4152, 4153-4154, 4155-4156, 4157-4158, 4159-4160, 4161-4162, 4163-4164, 4165-4166, 4167-4168, 4169-4170, 4171-4172, 4173-4174, 4175-4176, 4177-4178, 4179-4180, 4181-4182, 4183-4184, 4185-4186, 4187-4188, 4189-4190, 4191-4192, 4193-4194, 4195-4196, 4197-4198, 4199-4200, 4201-4202, 4203-4204, 4205-4206, 4207-4208, 4209-4210, 4211-4212, 4213-4214, 4215-4216, 4217-4218, 4219-4220, 4221-4222, 4223-4224, 4225-4226, 4227-4228, 4229-4230, 4231-4232, 4233-4234, 4235-4236, 4237-4238, 4239-4240, 4241-4242, 4243-4244, 4245-4246, 4247-4248, 4249-4250, 4251-4252, 4253-4254, 4255-4256, 4257-4258, 4259-4260, 4261-4262, 4263-4264, 4265-4266, 4267-4268, 4269-4270, 4271-4272, 4273-4274, 4275-4276, 4277-4278, 4279-4280, 4281-4282, 4283-4284, 4285-4286, 4287-4288, 4289-4290, 4291-4292, 4293-4294, 4295-4296, 4297-4298, 4299-4300, 4301-4302, 4303-4304, 4305-4306, 4307-4308, 4309-4310, 4311-4312, 4313-4314, 4315-4316, 4317-4318, 4319-4320, 4321-4322, 4323-4324, 4325-4326, 4327-4328, 4329-4330, 4331-4332, 4333-4334, 4335-4336, 4337-4338, 4339-4340, 4341-4342, 4343-4344, 4345-4346, 4347-4348, 4349-4350, 4351-4352, 4353-4354, 4355-4356, 4357-4358, 4359-4360, 4361-4362, 4363-4364, 4365-4366, 4367-4368, 4369-4370, 4371-4372, 4373-4374, 4375-4376, 4377-4378, 4379-4380, 4381-4382, 4383-4384, 4385-4386, 4387-4388, 4389-4390, 4391-4392, 4393-4394, 4395-4396, 4397-4398, 4399-4400, 4401-4402, 4403-4404, 4405-4406, 4407-4408, 4409-4410, 4411-4412, 4413-4414, 4415-4416, 4417-4418, 4419-4420, 4421-4422, 4423-4424, 4425-4426, 4427-4428, 4429-4430, 4431-4432, 4433-4434, 4435-4436, 4437-4438, 4439-4440, 4441-4442, 4443-4444, 4445-4446, 4447-4448, 4449-4450, 4451-4452, 4453-4454, 4455-4456, 4457-4458, 4459-4460, 4461-4462, 4463-4464, 4465-4466, 4467-4468, 4469-4470, 4471-4472, 4473-4474, 4475-4476, 4477-4478, 4479-4480, 4481-4482, 4483-4484, 4485-4486, 4487-4488, 4489-4490, 4491-4492, 4493-4494, 4495-4496, 4497-4498, 4499-4500, 4501-4502, 4503-4504, 4505-4506, 4507-4508, 4509-4510, 4511-4512, 4513-4514, 4515-4516, 4517-4518, 4519-4520, 4521-4522, 4523-4524, 4525-4526, 4527-4528, 4529-4530, 4531-4532, 4533-4534, 4535-4536, 4537-4538, 4539-4540, 4541-4542, 4543-4544, 4545-4546, 4547-4548, 4549-4550, 4551-4552, 4553-4554, 4555-4556, 4557-4558, 4559-4560, 4561-4562, 4563-4564, 4565-4566, 4567-4568, 4569-4570, 4571-4572, 4573-4574, 4575-4576, 4577-4578, 4579-4580, 4581-4582, 4583-4584, 4585-4586, 4587-4588, 4589-4590, 4591-4592, 4593-4594, 4595-4596, 4597-4598, 4599-4600, 4601-4602, 4603-4604, 4605-4606, 4607-4608, 4609

RESULT 15

T27935 hypothetical protein ZK617.lb - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C:Accession: T27935; T28031
R:White, S.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20442
A:Accession: T27935
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <MIL>
A:Cross-references: EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.lb
A:Experimental source: clone ZK617
R:Harris, B.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20458
A:Accession: T28031
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <WIZ>
A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.lb
A:Experimental source: clone ZK829
C:Genetics:
A:Gene: CESP:ZK617.lb
A:Map position: 4
A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 595/3
3067/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 5.8%; Score 402; DB 2; Length 7160;
Best Local Similarity 18.3%; Pred. No. 5e-11;
Matches 288; Conservative 153; Mismatches 532; Indels 600; Gaps 50;

QY	204	ERRRSSVPPIPADTRDNEVNHAEQ-QEAKRREAQAASEAAG-GDTTGGSPSSLIYY	261
Db	5470	EIRYGSTLMTVASDYNVREFTVDKLRNFEDYFRVAINAAGKGPLSPGPIKI--	5527
QY	262	EELPGQPRFTQKLSREVPGTGVQLDCIVVGIPPOVRWYCEGKELENSPDHIHQAG	321
Db	5528	QESGSRPQIVVKFEDTAQPNRRAVFTCEAVGRPEPTARWLNRGRELPESSRYRFEASD	5587
QY	322	NLHSLTAAFEEDTGRYSCFASNIYGDTSFA	354
Db	5588	GVYRFTKEVMDIDAGEYTVESNPGYSDTATANLVQAPPVIEKQVNTILPSGDLVRL	5647
QY	355	EIYIEGYS	366
Db	5648	KIYFSGTAPFRHSLVLRNEEDMDHPTIRIVFDDHILITIPALSVRAGRYEYTVSND	5707
QY	367	-----EGDP	370
Db	5708	GEATTGFWLNVTLGLPEAPQGPLHISNIGSPSTATLSWRPPTVDGGSKITSYVVEKRDLSKD	5767
QY	371	-----NKEEMNRI	400
Db	5768	EWVTVTNNVKDMNIVTGLFENHEHYEFVRSQAENGIGAPLVSEHPITARLPFDPPPTSP	5827
QY	401	AQHLV	422
Db	5828	NLEIVQGGDYVTLWSWRPQVLSGGGRULRGYIVVEKQEEHDEWFRCNQNPSPNNYVNP	5887
QY	423	-----LQGLDGKPIIAA	457
Db	5888	IDGRKYRVRFAVNDAGLSDLAELDQTLFQASGSGEGPKIVSPSLDNEEVGRCVTECE	5947
QY	458	VKGAPSKVEWYREGTLIEDSPDFRILOKPRSMAPPEEICTLVIAEVEFEDSCGCTCTA	517
Db	5948	ISGSPREYRMFGCKELAVDTSKYTLNKGDKQV-----LIINDLTSDDADEYTCRA	5999

Search completed: November 30, 2002, 19:07:41
Job time : 79 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2002, 18:51:30 ; Search time 18 Seconds
(without alignments)
3041.597 Million cell updates/sec

Title: US-09-818-990B-2

Perfect score: 6890

Sequence: 1 MQDDSIESTSISQLLRESY.....MESTWVYSCSRSVVESDEL 1320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	556.5	8.1	1906	1	KMLS_CHICK	P11799	gallus gall
2	546	7.9	1914	1	KMLS_HUMAN	Q15746	homo sapien
3	392	5.7	4393	1	PGBM_HUMAN	P98160	homo sapien
4	375.5	5.4	3707	1	PGBM_MOUSE	Q05793	mus musculus
5	364	5.3	1176	1	KMLS_BOVIN	Q28824	bos taurus
6	352.5	5.1	1147	1	KMLS_RABIT	P29294	oryctolagus
7	322	4.7	2481	1	UN52_CAEEL	Q06561	caenorhabdi
8	317	4.6	1465	1	MYM2_HUMAN	P54296	homo sapien
9	315	4.6	2012	1	DSCA_HUMAN	O60469	homo sapien
10	307	4.5	1447	1	DCC_MOUSE	P70211	mus musculus
11	301.5	4.4	1450	1	MP5F_CHICK	Q02173	gallus gall
12	300	4.4	1447	1	DCC_HUMAN	P43146	homo sapien
13	296.5	4.3	1709	1	SN_HUMAN	Q9B222	homo sapien
14	282.5	4.1	1666	1	MYM1_MOUSE	G62234	mus musculus
15	278	4.0	837	1	NCM2_MOUSE	O35136	mus musculus
16	272	3.9	725	1	NCA2_MOUSE	P13594	mus musculus
17	272	3.9	1115	1	NCA1_MOUSE	P13595	mus musculus
18	271	3.9	837	1	NCM2_HUMAN	O15394	homo sapien
19	271	3.9	853	1	NCA1_BOVIN	P31836	bos taurus
20	268.5	3.9	1070	1	PTK7_HUMAN	Q13308	homo sapien
21	267.5	3.9	1040	1	AXO1_HUMAN	Q02246	homo sapien
22	267	3.9	761	1	NCA2_HUMAN	P13592	homo sapien
23	264.5	3.8	848	1	NCA1_HUMAN	P13591	homo sapien
24	264.5	3.8	1036	1	AXO1_CHICK	P28685	gallus gall
25	264	3.8	1302	1	NRG_DROME	P20241	drosophila
26	263.5	3.8	1284	1	NRC_CHICK	P35331	gallus gall
27	261.5	3.8	1040	1	AXO1_RAT	P22063	rattus norv
28	261	3.8	858	1	NCA1_RAT	P13596	rattus norv
29	257.5	3.7	1694	1	SN_MOUSE	O62230	mus musculus
30	257	3.7	1493	1	NEO1_MOUSE	P97789	mus musculus
31	255	3.7	1461	1	NEO1_HUMAN	Q92859	homo sapien
32	253	3.7	1271	1	MYPC_CHICK	Q90688	gallus gall
33	248.5	3.6	1451	1	MYM1_HUMAN	P52179	homo sapien

RESULT 1				
KMLS_CHICK				
ID	KMLS_CHICK	STANDARD;	PRT;	1906 AA.
AC	P11799;	P19038;		
DT	01-OCT-1989	(Rel. 12, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Myosin light chain kinase, smooth muscle and non-muscle isozymes			
DE	(EC 2.7.1.117) (MLCK) [Contains: Telokin].			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A. (MLCK-210).			
RX	MEDLINE=96033976; PubMed=7589469;			
RA	Waterson D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,			
RA	Stepanova O.V., Shirinsky V.P.;			
RT	"Multiple gene products are produced from a novel protein kinase			
RT	transcription region.";			
RL	FEBS Lett. 373:217-220(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A. (MLCK-108).			
RX	MEDLINE=90192792; PubMed=2315320;			
RA	Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,			
RA	Means A.R.;			
RT	"Regulatory and structural motifs of chicken gizzard myosin light			
RT	chain kinase.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).			
RN	[3]			
RP	SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.			
RX	TISSUE=Fibroblast;			
RC	MEDLINE=90361738; PubMed=2202734;			
RA	Shoenaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,			
RA	Matrisian P.E., Guerra-Santos L., Wilson E., Lukas T.J.,			
RA	van Eldik L.J., Watterson D.M.;			
RT	"Use of DNA sequence and mutant analyses and antisense			
RT	oligonucleotides to examine the molecular basis of nonmuscle			
RT	myosin light chain kinase autoinhibition, calmodulin recognition, and			
RT	activity.";			
RL	J. Cell Biol. 111:1107-1125(1990).			
RN	[4]			
RP	SEQUENCE OF 1259-1906 FROM N.A.			
RX	TISSUE=Gizzard;			
RC	MEDLINE=87157587; PubMed=3030394;			
RA	Guerriero V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.;			
RT	"Domain organization of chicken gizzard myosin light chain kinase			
RT	deduced from a cloned cDNA.";			
RL	Biochemistry 25:8372-8381(1986).			
RN	[5]			
RP	SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).			
RC	TISSUE=Gizzard;			
RX	MEDLINE=93073972; PubMed=1444462;			
RA	Yoshikai S., Ikebe M.;			
RT	"Molecular cloning of the chicken gizzard telokin gene and cDNA.";			

P13590 gallus gall
Q90610 gallus gall
O63198 rattus norv
O70468 mus musculu
Q12860 homo sapien
O60885 homo sapien
P12960 mus musculu
P16170 xenopus lae
P70478 rattus norv
Q00872 homo sapien
P32004 homo sapien
P23468 homo sapien

ALIGNMENTS

Arch. Biochem. Biophys. 299:242-247(1992).
 [6]
 RN SEQUENCE OF 1750-1906 FROM N.A. (TELOKIN).
 RP MEDLINE=92236611; PubMed=1373815;
 RA Collinge M., Matrisian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
 van Eldik L.J., Watterson D.M.;
 RA "Structure and expression of a calcium-binding protein gene contained
 RT within a calmodulin-regulated protein kinase gene.";
 RL Mol. Cell. Biol. 12:2359-2371(1992).
 CC -1- FUNCTION: PHOSPHORYLATES A SPECIFIC SERINE IN THE N-TERMINUS OF A
 CC MYOSIN LIGHT CHAIN, WHICH LEADS TO THE FORMATION CALMODULIN/
 CC MLCK SIGNAL TRANSDUCTION COMPLEXES WHICH ALLOW SELECTIVE
 CC TRANSDUCTION OF CALCIUM SIGNALS.
 CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] -> ADP + [myosin
 CC light-chain] phosphate.
 CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms; MLCK-210/non-muscle
 CC form, MLCK-108/smooth-muscle form and telokin; are
 CC produced by alternative initiation.
 CC -1- TISSUE SPECIFICITY: TELOKIN IS EXPRESSED IN GIZZARD, HEART, LUNG,
 CC INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF THE
 CC EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE GIZZARD.
 CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52876; CAA37056.1; -
 DR EMBL: X52876; CAA37057.1; -
 DR EMBL: X52876; CAA37058.1; -
 DR EMBL: M31048; AAA49069.1; -
 DR EMBL: M14953; AAA69964.1; -
 DR EMBL: M96655; AAA49083.1; -
 DR EMBL: M88283; AAA48647.1; -
 DR EMBL: M88284; AAB53768.1; -
 DR PIR: A25810; A25810.
 DR PIR: A37099; A37099.
 DR PIR: S11652; S11652.
 DR HSP: P56276; 1TLK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00041; fn3; 1.
 DR Pfam: PF00047; ig; 9.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00060; FN3; 1.
 DR SMART: SM00410; IG_like; 1.
 DR SMART: SM00408; IGC2; 8.
 DR SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transferase: Serine/threonine-protein kinase; Calmodulin-binding;
 KW ATP-binding; phosphorylation; Immunoglobulin domain; Repeat;
 KW Alternative initiation.
 FT CHAIN 1 1906 MYOSIN LIGHT CHAIN KINASE, MLCK-210 (NON-
 FT MUSCLE ISOZYME).
 FT CHAIN 935 1906 MYOSIN LIGHT CHAIN KINASE, MLCK-108
 FT (SMOOTH-MUSCLE ISOZYME).
 FT CHAIN 1750 1906 TELOKIN.
 FT INIT_MET 935 935 FOR MLCK-108.
 FT INIT_MET 1750 1750 FOR TELOKIN.

FT	DOMAIN	1330	1400	FIBRONECTIN TYPE-III. PROTEIN KINASE. IG-LIKE C2-TYPE DOMAIN. ATP (BY SIMILARITY). ATP (BY SIMILARITY). BY SIMILARITY. CALMODULIN AUTOINHIBITION (AM13) REGION (POTENTIAL). CALMODULIN RECOGNITION (RS20) REGION (POTENTIAL). MOTIF IA. MOTIF IB. 4 X REPEATS, MOTIF IIA. IIA-1. IIA-2. IIA-3. IIA-4. 5 X REPEATS, MOTIF IIB. IIB-1. IIB-2. IIB-3. IIB-4. IIB-5. 4 X REPEATS, MOTIF III. III-1. III-2. III-3. III-4. CALMODULIN-BINDING. POLY-GLU. PHOSPHORYLATION. MOD_RES CONFLICT SEQUENCE	1453 1809 1459 1467 1482 1574 1716	1708 1876 1467 1482 1574 1728	1730	1749
FT	DOMAIN	1330	1400					
FT	DOMAIN	1453	1708					
FT	DOMAIN	1809	1876					
FT	NP_BIND	1459	1467					
FT	BINDING	1482	1482					
FT	ACT_SITE	1574	1574					
FT	DOMAIN	1716	1728					
FT	DOMAIN	1730	1749					
FT	DOMAIN	1317	1364					
FT	DOMAIN	1385	1402					
FT	DOMAIN	660	1833					
FT	REPEAT	660	676					
FT	REPEAT	758	774					
FT	REPEAT	1107	1123					
FT	REPEAT	1817	1833					
FT	DOMAIN	693	1866					
FT	REPEAT	693	708					
FT	REPEAT	791	807					
FT	REPEAT	1140	1156					
FT	REPEAT	1281	1297					
FT	REPEAT	1851	1866					
FT	DOMAIN	970	1226					
FT	REPEAT	970	987					
FT	REPEAT	999	1016					
FT	REPEAT	1061	1078					
FT	REPEAT	1209	1226					
FT	DOMAIN	1700	1763					
FT	DOMAIN	1896	1906					
FT	MOD_RES	1748	1748					
FT	MOD_RES	1762	1762					
FT	CONFLICT	1439	1439					
FT	SEQUENCE	1906 AA; 210445 MW; AD7D8A3B69E3363 CRC64;						
QY	Query Match	8.1%;	Score 556.5; DB 1; Length 1906;					
Db	Best Local Similarity	21.9%;	Pred. No. 5.8e-19;					
QY	Matches	237;	Conservative 151; Mismatches 438; Indels 255; Gaps 36;					
QY	267	QPPRFTQKLRSEVPEGTRVQVDCIVGPPQVRVYCEGKELENSPDTHIVQAG--NLH 324						
Db	26	EAPFTLPPRNIRVQIGATARFEGKVRGYEPEQITWYRNGHPLPEG-DHYVVDHSIRGIF 84						
QY	325	SLTIAEAFEDTGRYSCFASNIYGTDTSAEIVIEGVSSDSSEDPNKEEMRIQKPNEV 384						
Db	85	SLVKGVEGDSCKYTCEAANDGGVQVTVVELVEG-----NSLKYSLP 129						
QY	385	SSPPTTSVAVIPPAVQAQHLVAQPRVATIQCCOSPTNYLQGLDGKPII---AAPVTKML 441						
Db	130	SSAKTPGGRL--SVPPVEH-----RPSIWGESPPKFKATKP 162						
QY	442	QNLASASEGOLVPECRVKGAPSKVSEYREGTLIEDSPDFRIQKKPRSAEPPEICTLV 501						
Db	163	NRVVREGQTRGFSCKITGRPQVQVTKT-----GDHLQONERFNFEKTGQYLE 214						
QY	502	IAEVFAEDSGCFTCTASNKYGTGVSSIAQLHVRGNEDLSNNSLSHSANSTTNLAIEPQPS 561						
Db	215	IQNVQLADAGIYCTTVNSAGKASVSAELTVQG-----PDKT 251						
QY	562	PHSEPPSVGEQPKP-KLEGVLVNHNEPRSSSRIGIRVHP-----NLPEDDKSGEASSEA 615						
Db	252	DTHAQP--LCMPKPKPTLATKAIENSDFKQATNGIAKELKSTSTELMVETKDRLSAKKE 309						
QY	616	GVVTTTQTTPDSQERFNGOATKTPSPFPVKEPPVPLAKPLKLDSTQLOQLHNOVLEQH 675						
Db	310	TFYTSREAK-DGKQGO-NOEANAVPQESRGTKGPQV-----LQKTSITLQAV 357						
QY	676	QLQNPFPSPKPEPFXTMTVLNSNAPPVATTSXKQVKAPSSQTSFSLARPKYFFFTNTTAA 735						
Db	358	KAQPEKAEPPQ-----TTFIQAE-----DKRKTVPQLMTITT- 390						
QY	736	TVAPSSSPVFTLSSTPQTQRTVSKESLLVSHSPSVQTKSPGGL--SIQNEPLPPGGTPT 793						

```
Db 391 -----QENPSTGQVSPR-----SRETENRAGVRKSKVEKREP----- 424
QY 794 PPPTFTSIPSGNFOQPCVSPIPVSPTRSQNPVAFSLSSVLPSPALPPTNMXKLPSAP 853
Db 425 -----LGIPP--QFESR-----PQSLASGQELKFSKV-----SGKPPDV 460
QY 854 SMPSSQGLAKNTKSPQVNDNIR--ETKNVIRDLKGGKKTFTSDVRPNQOQYKISSFEQR 911
Db 461 EWFKEGVPIKTGEGIQIYEDGTHCLWLKACLDGSG-----SYSCAANPR 507
QY 912 LMNEIEPRL--ERTPVDESDEIQTGKCIAPTRFKLXHFRTVTEGSPVTFCKIV 969
Db 508 GQSTSWLLTVKRPKEE-----VAPCFSSVLKGTVSEGGDFVLQCYVG 552
QY 970 GIPPKVYWFKDGKQISKRNECHKMRREGDGTCSLHTESTTDDGNTYITMAANPOGRIS 1029
Db 553 GVPPEITWLLNPOIYAHSTFEA-----GVAKLTVDALPEDDGIYTCLEANNAGRAS 607
QY 1030 CSGLHMQSLPIRSLTSAGQSHRGRSRVQERDKEPLFRPHFLQAPCDMVHAGRL 1089
Db 608 CSAQVTVK-----EKKSSKAEGTQAALKNTKTFAPILFLGLTDLKVMDSQ 653
QY 1090 CRLDCKVSGLPPELWLLNQVPLPDASHKMLVRETGVHSLLDPLTQRDAGTYKCIAT 1149
Db 654 VIMTVEVSANPCPEIILWLNHNGKEI-QETEDFHEKKGNEYSLYIQEVPEDTKYTCEAW 712
QY 1150 NKTGONSFSLELSVAKEVKAPVILEKLNQCGVPEGHVPLRCRVIGMPPPVYKWKDN 1209
Db 713 NEIGETQATLVQEPDQGIQWFIKSPKSVTAAGQNVLIISCAIADGDFPTVHWPKDG 772
QY 1210 ETI-PCTRERISMHDPTGYACLIQPAKSDAGWYTLASNEAGIVSC-----T 1258
Db 773 QEITPGTGCIELQED---IFTLIRNVQSRHAGQYEIQLRNQVGECSCQVSLMLRESSA 829
QY 1259 ARDIYAQWHQHQPMPSPSSRYGSLKGLDIFSASFSSMESTWVYSCSSRVVVED 1318
Db 830 SRAEMLRDGRS--ASSGERRDGNGYALT-----FGRTSGFKKS---SSETRAABEEQ 878
QY 1319 E 1319
Db 879 E 879

RESULT 2
KMLS_HUMAN STANDARD: PRT: 1914 AA.
AC Q15746; Q14844; Q16794; Q95796; Q95797; Q95798; Q95799; Q9UBG5;
AC Q9UBI9;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin light chain kinase, smooth muscle and non-muscle isozymes
DE (EC:2.7.1.117) (MCKK) [Contains: Telokin (Kinase related protein)
DE (KRP)].
GN MYLK OR MCKK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=97304466; PubMed=9160829;
RA Garcia J.G.N., Lazar V.L., Gilbert-Mcclain L.I., Gallagher P.J.,
RA Varin A.D.;
RT "Myosin light chain kinase in endothelium: molecular cloning and
RT regulation.";
RL Am. J. Respir. Cell Mol. Biol. 16:489-494 (1997).
[2]
RP REVISIONS.
RA Birukov K.G., Garcia J.G.N.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
```

```
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).
RC TISSUE=Umbilical vein;
RX MEDLINE=99216419; PubMed=10198165;
RA Lazar V.L., Garcia J.G.N.;
RT "A single human myosin light chain kinase gene (MCKK; MYLK).";
RL Genomics 57:256-267 (1999).
[4]
RP REVISIONS (ISOFORM 2).
RA Birukov K.G., Garcia J.G.N.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 923-1914 FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=96121365; PubMed=8575746;
RA Potier M.-C., Chelot E., Pekarsky Y., Gardiner K., Rossier J.,
RA Turnell W.G.;
RT "The human myosin light chain kinase (MCKK) from hippocampus:
RT cloning, sequencing, expression, and localization to 3cen-q21.";
RL Genomics 29:562-570 (1995).
[6]
RP SEQUENCE OF 1614-1914 FROM N.A.
RC TISSUE=Lung, and Placenta;
RX MEDLINE=20007838; PubMed=10536370;
RA Watterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,
RA Shinsky V.P., Van Eldik L.J., Haiech J.;
RT "Analysis of the kinase-related protein gene found at human chromosome
RT 3q21 in a multi-gene cluster: organization, expression, alternative
RT splicing and polymorphic marker.";
RL J. Cell. Biochem. 75:481-491 (1999).
[7]
RP SEQUENCE OF 1456-1914 FROM N.A.
RC TISSUE=Placenta;
RA Watterson M.D.;
RA Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CALCULUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL
CC DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE
CC EDEMA FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO
CC CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE
CC AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED
CC BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT
CC IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC light-chain] phosphate.
CC -1- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; A NON-MUSCLE FORM (THE
CC LONGEST FORM), A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL
CC SECTION WITH NO CATALYTIC ACTIVITY); ARE PRODUCED BY ALTERNATIVE
CC INITIATION.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; 1 (SHOWN HERE), 2, 3A,
CC 3B, 4 AND DEL-1790; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE
CC EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN
CC CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE
CC NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
CC THE DOMINANT SPLICED VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
CC HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
CC -1- PTM: MCKK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U48959; AAC18423.2; -.
```

DR


```
Db 393 DVYSKAANRRIPMEGO-----RDSAPFKPESKQSEVKENQTVKRCVSGIPK---PE 444
Qy 916 IEFRLERTVP--DESDEIQHD-----EIPTGK----- 942
Db 445 VAWFLEGTVPVRQEGSIEYEDAGSHYLLKARFDSCTYCTASNAOGQVSCSWTLQV 504
Qy 943 -----IAPFDKRLKHFRTVTEGSPVFTCKIVGIPVKVYFKDQKQISKRNECHKMR 995
Db 505 ERLAVMEVAPSSVLLKDCAVIEGQDFVLQCSVRGTPVPRIITWLLNGQPIQYARSTCEA- 563
Qy 996 REGDGTCSLHISTSDDDGNTYIMANPQGRISCSGHLMVOSLPISRSLTSAGQSHRGR 1055
Db 564 ----GVAELHIQDALPEDGTGTYCLAENALQGVGSAAWTVH-----EKK 605
Qy 1056 SRVQERDKPELQERRPRPFELQAGDMVAHEGRCLCKDCKVSGSLPPELTWLLNGQPVLP 1115
Db 606 SRKSEVLLPVASKPTAPFLGLSLDKVMDGSOQVMTVQVSGNPPPEVWILHNGNEIQE 665
Qy 1116 DASHKMLVRETGVHSLLDPLTQRDAGTYKCIATNKTKGNSFSLSVVAKEYKAPVIL 1175
Db 666 SEDFHEQRTG--QHSLWIEQVFPEDTGTTCBAWNSAGEVRTQAVLTQVQEPHDGTQPWFI 724
Qy 1176 EKLONGCVPEGVHVRLECRVIGMPPPVFWKDKNETIPTCTRERISMQDTTGVACLLIQP 1235
Db 725 SKPRSVYASLGOSVLISCAIAGDPPTVHWRDGLKALCKDTGHEVFLQNEDEVFT-LVLKK 783
Qy 1236 AKKSDAGWYTLAKNAEAGIVSCTARDIYAQWHQIP---PMSVVRP-----SG 1281
Db 784 VQPHWAGQVEILLKNRVGCSQVSLMIONSSARALPRGREPASCEDLGGGVGADGGGS 843
Qy 1282 SRYGSL 1287
Db 844 DRYGSL 849

RESULT 3
PGBL_HUMAN
ID PGBL_HUMAN STANDARD; PRT: 4393 AA.
AC P98160; Q16287;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112994; PubMed=1730768;
RA Kallunki P., tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Skin, and Colon;
RC MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 1018-1472 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
```

```
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [4]
RP SEQUENCE OF 892-1398 FROM N.A.
RX TISSUE=Fibrosarcoma;
RC MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [5]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC -!- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62515; CAA44373.1; -
DR EMBL; M85289; AA52700.1; -
DR EMBL; M64283; AA52699.1; -
DR EMBL; S76436; AA321121.2; -
DR EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR HSPG; P00740; 1EDM.
DR Siena-2DPAGE; P98160; -
DR Genew; HGNC:5273; HSPG2.
DR MIM; 142461; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00047; Ig; 22.
DR Pfam; PF00052; laminin_B; 3.
DR Pfam; PF00053; laminin_EGF; 7.
```



```
Db 3349 --NELLHFERAAPDSGR-----YCRVTNKNVGSAAFAQLLVGGPGSLPAT----- 3394
QY 932 IQHDEIPTGKC-----IAPIDFKRLKHFRTVEGSPVTCTKIVGIPVKVWFKDGKOISK 987
Db 3395 -----SIPAGSTPTVQVTPQLETK-----SIGASVEFHCAVPDRGTQLRWFKEGGGLPP 3444
QY 988 RNECHKMRREGDGTCSLHIESTTSDDDGNTYIMANPOGRISCSGHLMVQSLPIRSRLTS 1047
Db 3445 GHS-----VQDGV--LRIONLQSCQGTVCQAHPGWAQAQAQLVQALP--SVLIN 3494
QY 1048 AGQSHRGSRVQERKDEPLQERFRPHFLQAPGDVWAHEGLRCLDKCVSLPPLPETWL 1107
Db 3495 I-----RTSVQ-----TVVGHAVEFECLALGDPRKQVTVWS 3525
QY 1108 LNGOPVLPSHAKMLVRETGVHSLIDPLTQRDAGTKYKCIATNKTGNSFSLSVVAKE 1167
Db 3526 KVGGLHRLPG-----IVSGGV--VRIAHVELADAGVRCRTATNAAGTQSHVLLVQALP 3578
QY 1168 VKRAPVILEKLQNGVPEHPVRLCEVRIGMPPEVFWYKK-DNETIPTCTRERISMHQDTT 1226
Db 3579 QISMP-----QEVRRVPAGSAVPPCIASGVPTPDISWSKLDGSLPPDSRLNNM----- 3627
QY 1227 GYACLLIOPAKKSDAGWYTSIAKNEAGIVSCTARLDI 1263
Db 3628 ----LMLPSVQPDAGTYVCTATNRQGVKRAFAHLQV 3660

RESULT 4
PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Melanoma.
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule."
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RS SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan."
RL J. Biol. Chem. 263:16379-16387(1988).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
```

```
FT DOMAIN 1126 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 1158 LAMININ EGF-LIKE 6.
FT DOMAIN 1208 LAMININ EGF-LIKE 7.
FT DOMAIN 1265 LAMININ EGF-LIKE 8.
FT DOMAIN 1324 LAMININ EGF-LIKE 9.
FT DOMAIN 1334 LAMININ EGF-LIKE 9 (N-TERMINAL).
FT DOMAIN 1335 LAMININ DOMAIN IV 3 (DOMAIN III C).
FT DOMAIN 1530 LAMININ EGF-LIKE 9 (C-TERMINAL).
FT DOMAIN 1562 LAMININ EGF-LIKE 10.
FT DOMAIN 1563 LAMININ EGF-LIKE 11.
FT DOMAIN 1612 LAMININ EGF-LIKE 11.
FT DOMAIN 1677 1771 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 1677 1771 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 1772 1865 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 1866 1954 IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN 1955 2049 IG-LIKE C2-TYPE DOMAIN 6.
FT DOMAIN 2050 2148 IG-LIKE C2-TYPE DOMAIN 7.
FT DOMAIN 2149 2244 IG-LIKE C2-TYPE DOMAIN 8.
FT DOMAIN 2245 2343 IG-LIKE C2-TYPE DOMAIN 9.
FT DOMAIN 2344 2436 IG-LIKE C2-TYPE DOMAIN 10.
FT DOMAIN 2437 2532 IG-LIKE C2-TYPE DOMAIN 11.
FT DOMAIN 2533 2619 IG-LIKE C2-TYPE DOMAIN 12.
FT DOMAIN 2620 2721 IG-LIKE C2-TYPE DOMAIN 13.
FT DOMAIN 2721 2809 IG-LIKE C2-TYPE DOMAIN 14.
FT DOMAIN 2810 2895 IG-LIKE C2-TYPE DOMAIN 15.
FT DOMAIN 2896 2980 LAMININ G-LIKE 1.
FT DOMAIN 2984 3162 EGF-LIKE.
FT DOMAIN 3163 3241 LAMININ G-LIKE 2.
FT DOMAIN 3245 3425 LAMININ G-LIKE 3.
FT DOMAIN 3518 3705 HEPARAN SULFATE (POTENTIAL).
FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
FT SITE 76 78 HEPARAN SULFATE (POTENTIAL).
FT SITE 3615 3617 MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 342 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 428 479 BY SIMILARITY.
FT DISULFID 764 773 BY SIMILARITY.
FT DISULFID 766 780 BY SIMILARITY.
FT DISULFID 783 792 BY SIMILARITY.
FT DISULFID 795 811 BY SIMILARITY.
FT DISULFID 814 829 BY SIMILARITY.
FT DISULFID 816 839 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 869 BY SIMILARITY.
FT DISULFID 1159 1168 BY SIMILARITY.
FT DISULFID 1161 1175 BY SIMILARITY.
FT DISULFID 1178 1187 BY SIMILARITY.
FT DISULFID 1190 1206 BY SIMILARITY.
FT DISULFID 1209 1224 BY SIMILARITY.
FT DISULFID 1211 1234 BY SIMILARITY.
FT DISULFID 1237 1246 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1275 1287 BY SIMILARITY.
FT DISULFID 1277 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1307 1322 BY SIMILARITY.
FT DISULFID 1363 1372 BY SIMILARITY.
FT DISULFID 1563 1572 BY SIMILARITY.
FT DISULFID 1565 1579 BY SIMILARITY.
FT DISULFID 1582 1591 BY SIMILARITY.
FT DISULFID 1594 1610 BY SIMILARITY.
FT DISULFID 1613 1628 BY SIMILARITY.
FT DISULFID 1615 1638 BY SIMILARITY.
FT DISULFID 1641 1650 BY SIMILARITY.
FT DISULFID 1653 1668 BY SIMILARITY.
```

```
FT DISULFID 1792 1839 BY SIMILARITY.
FT DISULFID 1886 1932 BY SIMILARITY.
FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.
FT DISULFID 2365 2413 BY SIMILARITY.
FT DISULFID 2456 2506 BY SIMILARITY.
FT DISULFID 2554 2599 BY SIMILARITY.
FT DISULFID 2641 2686 BY SIMILARITY.

Query Match 5.4%; Score 375.5; DB 1; Length 3707;
Best Local Similarity 19.5%; Pred. No. 4.3e-10;
Matches 280; Conservative 160; Mismatches 491; Indels 507; Gaps 62;

QY 26 HRGNERSRAEPSPNCH-----FGSPSGA-----AEGGGQDDLPDLAPLSQE 70
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1950 HGGGPRVQVSPERTQHEGRTVRLYCRAGVPSASITWRKEG-----SLP----- 1996

QY 71 ELDESVNRLARLAINVDPLEKADETOARKRLSPDQMKHSPNLSFEPNFCQDNPRPTSKE 130
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1997 -FRQAHGSRRLRLHMSVADSGEYVCRANNIDAQETSIMISVSP-----STNS 2044

QY 131 SPQAKRPQVCSETQSKKVFLNKAADFIEELUSLFKSHSSKRI----RPRACKNHKSKLE 186
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2045 PPAPASAPAPIRIESSSRVABEGQTLD---LNCVVPGHAHAQVTHKRGSLPHTHQTG 2100

QY 187 SQNKVMENSSFSDLSERRERSVPIPADTRDNEVNHAEQOEAKRREAQAAS--E 244
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2101 SRLRYQVSSADSGEYCVLSSSGPL-----EASVLVSITP 2137

QY 245 AAGDTPPGSSPSSLYEPLGQPPRTQKLRSREVPEGTQVLDLCIVVGIPPVQVRWYC 304
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2138 AANVHIPGVVP-----PIRI--ETSSSRVABEGQTLDSLSCVPGQAHQVTHK 2184

QY 305 EGKELENSPDHIVQAGNLHSLTIAEAFEDTGRYCSFASNIYTDSTA----- 354
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2185 RGGSLPAGHVH-----GHMLRNRVSPADSGEYSCQVTGSSGTLEASVLVTIEASEPS 2238

QY 355 -----EYIEGVSSDSEGD-----PNKEEMN----- 376
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2239 PIPAPGLAQPVIIESSSSHLTEGQTVDLKCVVPGQAHQVTHKRGSLPARHGTGSL 2298

QY 377 RIQKPNEV-----SSPPTTSVIPPVPAQAH-----LVAQPRVATIQOC 416
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2299 RLYQLSPADSGEYVCQVAGSSHPHEASFKLTVPSQNSSSFRSLRSPVISIEPPSSTVQQ- 2357

QY 417 QSPTNYLQGLDGKPIIAAPVFTKMLQNLASEGQLVVFECRV-KGAPSPKVEWVREGTLI 475
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2358 -----GQDASFCKLIHEGAMPKIKVENKIRDOEL 2385

QY 476 ED---SPD---FRILOKKPRMAEPEICTLVIAEVAEDSGCFTCTASKNYCTVSSIA 528
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2386 EDNVHISPNIGSIITIVAGPATM-EPT-----ACVASNVYMAQSVV 2426

QY 529 QLHVRGNEDLS--NNGSLH-SANSTTNLAAT---EPQSPPHSE---PPSVEQPPKPKLE 579
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2427 NLSVHGPTTVSVLEGGPVHVKGMDITLECISSGEPSSSPRWTRLGIPVKLE----PRMF 2482

QY 580 GVLVNHNEPRSSRIGRLRVHFNLPEDDKGSPASSEAGVVTTRTPRPSXOERFNGQATKT 639
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2483 GLMNSHAMLKIAS-----VKPSDAGTVVCAQNALGTAQKQVELIVD-----TGT 2527

QY 640 PEPSPVKEPPPVLA KPCLDSTQLQHLNQVLLQHQHQLQ-----NPPSSPKPEPFX 691
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2528 VAPGTP-----QVQVESELTLLEAGHTATLHCSATGNPPP----- 2562

QY 692 MVLNSNAPPVATTSXQVKAP-----SSQTESLAR-----PKYFFPSTNTAATVA 738
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2563 -----THWSKLRAPLWQHRIEGNTLVIIPRVAQODSGQYICNATNSAGHTEA 2610

QY 739 PSSSPVFTLSTPQTQIORTVSKESLLVSHPSVQTKSPGGLSIQNEPLPPGTEPTPPPT 798
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
Db 2611 ---TVVLHVESPYPY-----TIPEHTSAO-----PGML-VOLQCLAHG-----TPPLT 2650
QY 799 FSIPGNQFQRCVSPIPVSPTRSQINQVAFLLSSVPSLPAPPTNAMKLPRSAQSMPSQ 858
Db 2651 YQ-----NSLVGGVLPKAVV-----RNQLLRLEPTVP-- 2678
QY 859 GLAKKNTKSPQVNDNIR---ETKNVIRDLGKKITFSDVRPNQOEYKISSFEQRLMNE 915
Db 2679 -----EDSGRYRCQVSNRV---GSAEFAQVL----- 2702
QY 916 IEPRLERTPVDESDDEIHDIEPTGK---IAPIDKRLKHPRVTGEGSPVPTCKIVGI 971
Db 2703 -----VQSSSNLPTDTPSGGSTPTVQVTPQLETR-----NIGASVEFHCAVPNE 2747
QY 972 PVPKVYWFKDGKQISKRNEHCKMRREGDGTCSLHIESTTSDDGNTTMAANPQGRISCS 1031
Db 2748 RGTHLRLKEGGLPPGHS-----VQDGV--LRIONLQNCOGTYVCOAHGPGWQAQAT 2799
QY 1032 GHLMVQSLP-----ISRRLTSAGQSHRGRSRVQERKEPLQERFRPHFLQAPGDMVAHE 1086
Db 2800 AQLIVQALPSVLINVRTSVHSV-----VVGHS 2826
QY 1087 GLRLCDLCKVSGLPPELTWLLNGQVLPDASHKMLVRETGVHSLLDPLTQRDAGTYKC 1146
Db 2827 ---VEFECLALGDKPKQVITWSKVGGLRPG-----IVQSGT-IIIRIAHVELADAGQYRC 2876
QY 1147 IATNKTGQNSFSLSELSVA--KEVKKAPVILEKLQNGVPEGHVRLCECRVIGMPPPVFVW 1205
Db 2877 AATNAAGTTQSHVLLVQALPQISTPPEI-----RVFAGSRAVPFCMASGYPTTAIW 2929
QY 1206 KK-DNETICTRERISMHOQDTTGYACLLIOPAKKSDAGWYTTLSAKNAGIVSCTARLDI- 1263
Db 2930 SKVDGDLPPDSRLNNN-----LMLPSVRPEDAGTVCTATNRQGVKAFAYLQVP 2980
QY 1264 -----YAQHHHQIOPPMVSRVSGSYGSLTSKGLDIFSAF--SSMESTWVYSCSRS 1313
Db 2981 ERVPIYPTQPSYFLPLPTIKDAYRKF-----EIKITFRPDSADGMLLYNQKRS 3030

RESULT 5
KMLS_BOVIN
ID KMLS_BOVIN STANDARD; PRT: 1176 AA.
AC Q28824;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin light chain kinase, smooth muscle (EC 2.7.1.117) (MLCK)
DE [Contains: Telokin].
GN MLK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Stomach;
RX MEDLINE=93203148; PubMed=1284247;
RA Kobayashi H., Inoue A., Mikawa T., Kuwayama H., Hotta Y., Masaki T.,
RA Ebashi S.;
RT "Isolation of cDNA for bovine stomach 155 kDa protein exhibiting
RT myosin light chain kinase activity.";
RL J. Biochem. 112:786-791(1992).
CC -1- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS.
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC light-chain] phosphate.
CC -1- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS OF MLCK ARE PRODUCED BY
CC ALTERNATIVE INITIATION, A NON-MUSCLE FORM (WHICH IS THE LONGEST
CC FORM); A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL SECTION WITH
```

```
CC NO CATALYTIC ACTIVITY).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S571131; AAB25794.1; -.
CC HSP; P56276; 1TLK.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR003598; Ig_c2.
CC InterPro: IPR003600; Ig_like.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam: PF00041; fn3; 1.
CC Pfam: PF00047; ig; 2.
CC Pfam: PF00069; pkinase; 1.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00060; FN3; 1.
CC SMART: SM00408; IGC2; 2.
CC SMART: SM00220; S_TKC; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
CC Transferase: Serine/threonine-protein kinase; Calmodulin-binding;
CC ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
CC Alternative initiation.
CC CHAIN 1 1176 MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
CC ISOZYME.
CC INIT_MET 1022 1176
CC DOMAIN 100 291 FOR TELOKIN.
CC REPEAT 100 111 16 X 12 AA TANDEM REPEATS.
CC REPEAT 112 123 2.
CC REPEAT 124 135 3 (INCOMPLETE).
CC REPEAT 136 147 4.
CC REPEAT 148 159 5.
CC REPEAT 160 171 6.
CC REPEAT 172 183 7.
CC REPEAT 184 195 8.
CC REPEAT 196 207 9.
CC REPEAT 208 219 10.
CC REPEAT 220 231 11.
CC REPEAT 232 243 12.
CC REPEAT 244 255 13.
CC REPEAT 256 267 14.
CC REPEAT 268 279 15.
CC REPEAT 280 291 16.
CC DOMAIN 603 673 FIBRONECTIN TYPE-III.
CC DOMAIN 725 980 PROTEIN KINASE.
CC DOMAIN 972 1035 CALMODULIN-BINDING.
CC NP_BIND 1084 1151 IG-LIKE C2-TYPE DOMAIN.
CC BINDING 731 739 ATP (BY SIMILARITY).
CC ACT_SITE 754 754 ATP (BY SIMILARITY).
CC ACT_SITE 846 846 BY SIMILARITY.
CC DOMAIN 1171 1176 POLY-GLU.
CC SEQUENCE 1176 AA; 128824 MW; F53DC6D4D42D4B97 CRC64;

Query Match 5.3%; Score 364; DB 1; Length 1176;
Best Local Similarity 19.8%; Pred. No. 3 9e-10;
Matches 238; Conservative 159; Mismatches 495; Indels 310; Gaps 45;

QY 22 AET-RHRGNRNRAPSSNPCHFGSPSGAAEGGGODDLPDLISAFUSQELDESVDLAR 80
Db 122 AETPKPLGN-----VKPAETPKPLGSTKPA-----ETPKPLGSTKPAETPKLGNVK 168
```


CC -1- DEVELOPMENTAL STAGE: SYNTHESIZED EARLY IN EMBRYOGENESIS.
CC -1- SIMILARITY: CONTAINS 3 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 16 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -1- SIMILARITY: TO PERLECAN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L13458; AAA28156.1; -
DR HSPF; P01130; ILDR.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR000034; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000519; P_trefol.
DR Pfam: PF00047; Ig; 16.
DR Pfam: PF00052; laminin_B; 2.
DR Pfam: PF00053; laminin_EGF; 5.
DR Pfam: PF00057; ldl_recept_a; 3.
DR ProDom: PD003031; Laminin_B; 2.
DR SMART: SM00180; EGF_Lam; 5.
DR SMART: SM00410; IG_Like; 5.
DR SMART: SM00408; IGC2; 11.
DR SMART: SM00192; LDLa; 3.
DR SMART: SM00281; Lamb; 2.
DR SMART: SM00018; P; PARTIAL.
DR PROSITE: PS00022; EGF_1; 4.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01209; LDLRA_1; 3.
DR PROSITE: PS00068; LDLRA_2; 3.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 7.
KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
KW Basement membrane; Extracellular matrix; Alternative splicing;
KW Laminin EGF-like domain.
FT SIGNAL 1 18
FT CHAIN 19 2481
FT DOMAIN 19 148
FT DOMAIN 149 188
FT DOMAIN 190 229
FT DOMAIN 233 272
FT DOMAIN 273 363
FT DOMAIN 384 431
FT DOMAIN 432 441
FT DOMAIN 442 632
FT DOMAIN 633 665
FT DOMAIN 673 719
FT DOMAIN 720 729
FT DOMAIN 730 920
FT DOMAIN 921 953
FT DOMAIN 954 1003
FT DOMAIN 1010 1059
FT DOMAIN 1060 1110
FT DOMAIN 1133 1216
FT DOMAIN 1227 1311
FT DOMAIN 1318 1403
FT DOMAIN 1415 1499
FT DOMAIN 1507 1592
FT DOMAIN 1598 1682
FT DOMAIN 1699 1785
FT DOMAIN 1794 1878
FT DOMAIN 1887 1976
FT DOMAIN 1977 2070
FT DOMAIN 2078 2162

FT DOMAIN 2174 2260
FT DOMAIN 2267 2347
FT DOMAIN 2354 2436
FT DISULFID 66 114
FT DISULFID 149 161
FT DISULFID 156 174
FT DISULFID 168 183
FT DISULFID 190 202
FT DISULFID 197 215
FT DISULFID 209 224
FT DISULFID 233 246
FT DISULFID 240 259
FT DISULFID 253 268
FT DISULFID 954 963
FT DISULFID 956 970
FT DISULFID 973 982
FT DISULFID 985 1001
FT DISULFID 1010 1020
FT DISULFID 1012 1026
FT DISULFID 1029 1038
FT DISULFID 1041 1057
FT DISULFID 1060 1068
FT DISULFID 1062 1078
FT DISULFID 1081 1090
FT DISULFID 1093 1108
FT DISULFID 1151 1199
FT DISULFID 1337 1383
FT DISULFID 1434 1480
FT DISULFID 1526 1572
FT DISULFID 1617 1662
FT DISULFID 1718 1766
FT DISULFID 1813 1860
FT DISULFID 1906 1953
FT DISULFID 1997 2052
FT DISULFID 2098 2146
FT DISULFID 2194 2241
FT DISULFID 2283 2328
FT DISULFID 2373 2419
FT CARBOHYD 1421 1421
SQ SEQUENCE 2481 AA; 270812 MW; E47FBE2B310F5031 CRC64;
Query Match 4.7%; Score 322; DB 1; Length 2481;
Best Local Similarity 21.2%; Pred. No. 8.8e-08;
Matches 248; Conservative 144; Mismatches 458; Indels 318; Gaps 56;
QY 263 EPLQOPRFTQKLSREVPECTRVQLDCIVVGIIPPQVWRWCEKLENSPDIIH-IVQAG 321
DB 1226 QPVIDPPLHV-----VNEGEPAAFRCWVPGIPDCQITWH---REQLGGPLPHGVYQTG 1275
QY 322 NLHSLTAEAFEDTGRVSCFASNIYGTDTSAEIIYIEGVSSSDSGDPNKEENRIQKP 381
DB 1276 N--ALKTPQSOLHAGRYICSAANQYGTGSPPAVLEVKRPVDPKPIQTVDRDQPA 1333
QY 382 NEVSPPTTS-----AVTPPAVQAHLVAQPRVATIQCCQSPNTYLOGLDGKP 430
DB 1334 RFKCWPCGNSNVQLRWSRPGGAPLPSGVQEQGILHITPRASDQEVGVVCTATDPSDNT 1393
QY 431 IIAAPVTMKLQ-----NLSASEQLVVFECRVKGAQSPKVEWYREGTLI 475
DB 1394 LQSEPVQLNRDPAAPPQGAAPQIDPPNQTNVNDPAQFCRWPGQRAQLKWSRK----- 1449
QY 476 EDSPDFRILQKKRSMAPPEICTLVIAEVEAFEDSGCTCTASNKYGT--VSSTAQLHVR 533
DB 1450 ----DGRPL---PNGILERDGF--LRIDKSQLHDAGEVECTSTPDGSLQSPARLVN 1500
QY 534 GNEDLSNNGSLHSANSTNLAAIEPQPSPP-----HSEPPSVE--QPPKPKLEGVLVNH- 585
DB 1501 QPQ-----AIQPVDPVPQVTVNEGEPSRIRCWVPGHPNIQLQFVKRG 1542
QY 586 NEPRSSSRIGLRVHFNLPEDDKGSE-----ASSEAGVVTTRQTRP--DS 627
DB 1543 RRLPAHARFSQGNLEIPRTLKSDSEDEYICATDPTTNRNPVSNPARVIVKSPIRPIDP 1602

Db 327 -----KWKFFGEGQASLSFSLHKLDEGLYTLRIVRSGGVTDHSAFLFVRDADPLVT 380
Qy 541 NG-----SLRSANSTNLAIAEIPQSPHSEPPSEOPKPKLEGVLVNHNEPRSSR 593
Db 381 GAPGAPDQLCHDANRDYVITWKP-----PNTTTESP---VMGYFVDRCEVGTNNW 429
Qy 594 IGLRVHFNLPEDDKGSEASSAGVYVTTTQTRPDSXQER---FNGQATKTPSPSPVKEPP 650
Db 430 V-----QCNDAPVKICKYPTVGLFEGRSYIFRVRVANSAGISRPSRSD 473
Qy 651 PVLAKPKLDSTQLQOLH-----NOVLLQHQLOL-----NPP-----P 682
Db 474 AVAALDPLDLRLQAVHLEGEKEIAIYODDLEGDAQVPGPTGVHASEISRNYVVLWSEP 533
Qy 683 SSPKEFPXMTVLNSAPPVYV---TSXKOVKAPSSQTFESLARPK-YFF-----727
Db 534 PTPRGKDLMYFIEKSVVSGSTWVRVNAQTAVRPRVAFDLMGCKSVFVRLVANSRHGL 593
Qy 728 --PSTNTTA-----ATVAPSSPVFTLSS-----TPQTIQ---RTVSKESLLVSHPSVQTKS 774
Db 594 SEPSEITSPQAQDVTVVP-SAPGRVLASRNTKTSVVVQWDRPKHEEDLLGYVYDC---C 649
Qy 775 PGLLSIQNEPLPGPTPTPPP---TFSPISQNG---FQPRCVSPDPVSPTSRIQNPVA 828
Db 650 VAGTNLWE-----PCNHKPIGYNRFVYVHGLTTGEQYIFRVKAVNAVGMSENSQESDVIK 703
Qy 829 FLSSV-LPSLP-----AIP-----PTNAXMLPR-----SA 852
Db 704 VQALTVPSHPYGTLLNCDGHSMTLWKVKPKFSGGSPILGYLDKREVVHKNHNEVNSS 763
Qy 853 PSMPs-----OGLAKNNTKS-----POPVND---DN 875
Db 764 PSKPTILTVDGLTBSLYEFKIAAVNLAGICEPDSPEHFKCEAWTPEPGPAYDLTFCE 823
Qy 876 IRETKNAVI-----RDLGKKTTFSVDRNQOBYKTSFEQRLMN 914
Db 824 VRDTSVLMLKAPVYSGSSPVGYFVDFREDAGEWITVDOTTTASRYLKVSDLQQG---K 881
Qy 915 EIEPRLERT-----PVDESDDIEQHDIEPTGKCIAP-----FDKRLKHFRVTE 958
Db 882 TYVFRVAVNANGVKGSFSDTSEPVLVAPGCTKEISAGVDQGNIIYLGFCQ-----EMTD 937
Qy 959 GSPVFTTKIVGIPVPKVKDGKQISKRNECHKMRREGDGTCSLHIESTSDDDGNYT 1018
Db 938 ASQFT-----WCKSVYEEISD-DEREKIETVGDHS-KLVLNKPKDKEDLGTYS 981
Qy 1019 IMAANPOG-----RISCSGHLMVQSLPIRSRL-----1045
Db 982 VSVSDTDGVSSSFVLDPEELERLMAISNEIKNPTIPLKSELAYEIFDKGRVRFWLQAEHL 1041
Qy 1046 -----TSAGQSHRGRSRVQ-1059
Db 1042 SPDASYRFIINDREVSDSEIHRIKCDKATGIIEMWDRFSTENGSTYTVVTHDGAKSQS 1101
Qy 1060 -----ERDKPELQE-RFRFRPHFLQAPDMVAH-----EGRLCRLDCKVSGLPPEL-1104
Db 1102 SLVLIGDAFKTVLEAEFORKEFLKQGPFAEYLVHMDVTBECEVRVCKVANTKKETVF 1161
Qy 1105 TWL-----LNGQVPLDASHKMLVRETGVHSLIIDPLTQRDAGTYKCIATNKTGNSPSLE 1160
Db 1162 KWLKDALYETETLPNL-----ERGICELLIPKLSKKHGDGEYKATLKDRGDGVDSILE 1214
Qy 1161 LS-----VVAKEVKKAPVI-----1174
Db 1215 IAGKYDDMILASRVCKSAPLKVLCCTPGIRLQCFMKYFTDEMKNVCHCKDAKISS 1274
Qy 1175 -----LEKIQNC-----1181
Db 1275 EHRIGGSEEMALQICEPTKDKGKYTFEYFDGKDNHQRSLDLSGQAFDAFAEFQOFK 1334
Qy 1182 -----GVP-----EGHPVRLCECRVIGMPPPVYVYKKNQNETIPTCERISM 1221
Db 1182 -----GVP-----EGHPVRLCECRVIGMPPPVYVYKKNQNETIPTCERISM 1221

Db 1335 AAAPFAENRGRGLGGLPDVVTIMEGKTLNLTCTVFGNPDPEVIFWFKNDQDITOLS-EHFSV 1393
Qy 1222 HQDTTGYACLLIQAPKAKSDAGWYTLAKNEAGIVSCSTARLDIYAOWHQIP---PPMSVR 1278
Db 1394 KVEQAKYVMTIKGTVSEDSKYSINIKNKYGGKIDVTSVY-KHGEKIPDMAPPOQAK 1452
Qy 1279 P 1279
Db 1453 P 1453
RESULT 9
DSCA_HUMAN
ID DSCA_HUMAN STANDARD; PRT; 2012 AA.
AC O60469; O60468;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Down syndrome cell adhesion molecule precursor (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Huot Y.-K., Haendel M.A., Hubert R., Chen X.-N.,
RA Lyons G.E., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system";
RL Hum. Mol. Genet. 7:227-237(1998).
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=20384934; PubMed=10925149;
RA Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21";
RL Nature 405:311-319(2000).
CC
CC -1- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM/CHD2-52 (SHOWN HERE)
CC AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC
CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC
CC -1- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF023450; AAC17967.1; -
EMBL; AF023449; AAC17966.1; -
EMBL; AF217525; AAF27525.1; -
EMBL; AL163283; CAB90464.1; -
EMBL; AL163282; CAB90436.1; -
EMBL; AL163281; CAB90444.1; -
Genew; HGNC:3039; DSCAM.
MIM; 602523; -
InterPro; IPR003961; FN_III.
InterPro; IPR003962; FN_III.Repeat.
InterPro; IPR003006; Ig_MHC.
InterPro; IPR003598; Ig_C2.
InterPro; IPR003600; Ig_Like.
Pfam; PF00041; fn3; 6.
PRINTS; PRO0014; ENTPEI11.
SMART; SM00060; FN3; 6.
SMART; SM00410; Ig_Like; 2.
SMART; SM00408; IgC2; 7.
Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
Transmembrane; Alternative splicing.
FT SIGNAL 1 17 POTENTIAL
FT CHAIN 18 2012 DOWN SYNDROME CELL ADHESION MOLECULE.
FT DOMAIN 18 1595 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1596 1616 POTENTIAL.
FT DOMAIN 1617 2012 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 39 109 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 138 204 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 239 300 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 328 392 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 421 491 IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN 518 582 IG-LIKE C2-TYPE DOMAIN 6.
FT DOMAIN 610 676 IG-LIKE C2-TYPE DOMAIN 7.
FT DOMAIN 704 773 IG-LIKE C2-TYPE DOMAIN 8.
FT DOMAIN 802 872 IG-LIKE C2-TYPE DOMAIN 9.
FT DOMAIN 885 972 FIBRONECTIN TYPE-III 1.
FT DOMAIN 984 1076 FIBRONECTIN TYPE-III 2.
FT DOMAIN 1088 1177 FIBRONECTIN TYPE-III 3.
FT DOMAIN 1189 1273 FIBRONECTIN TYPE-III 4.
FT DOMAIN 1300 1366 IG-LIKE C2-TYPE DOMAIN 10.
FT DOMAIN 1380 1463 FIBRONECTIN TYPE-III 5.
FT DOMAIN 1477 1562 FIBRONECTIN TYPE-III 6.
FT DISULFID 46 102 BY SIMILARITY.
FT DISULFID 145 197 BY SIMILARITY.
FT DISULFID 246 293 BY SIMILARITY.
FT DISULFID 335 385 BY SIMILARITY.
FT DISULFID 428 484 BY SIMILARITY.
FT DISULFID 525 575 BY SIMILARITY.
FT DISULFID 617 669 BY SIMILARITY.
FT DISULFID 711 766 BY SIMILARITY.
FT DISULFID 809 865 BY SIMILARITY.
FT DISULFID 1307 1359 BY SIMILARITY.
FT CARBOHYD 28 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 795 795 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1142 1142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1250 1250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1271 1271 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1488 1488 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1562 1571 NFATLNDGS -> KEARCKES (IN SHORT ISOFORM).
FT VARSPLIC 1572 2012 MISSING (IN SHORT ISOFORM).
FT CONFLICT 1893 2012 HRPGLDHLPLPYLRMDFFLNRRGGTSDLSLQACLEPOK
FT FT SRTLKRPVLEPTPEMAASASSTREGQWQGVATLPOR
FT FT EGAELOAAKMSQSSQSLSDSRGHLKNNPYAKSYTLV ->
FT FT IGQVTSYICLIHLEWTF (IN REF. 1).
SQ SEQUENCE 2012 AA; 222259 MW; 0E33CFB781A08334 CRC64;
Query Match 4.6%; Score 315; DB 1; Length 2012;
Best Local Similarity 18.2%; Pred. No. 1.5e-07;
Matches 212; Conservative 163; Mismatches 424; Indels 364; Gaps 48;
QY 257 SLLYYEPLQPPRFTOKLSRVPECTRVQLDCIVVGIIPPQVRWY-CEGKELENSPDI 315
DB 22 SLLYFNASLQEVVFAST-----TGLVPCPAAGIPPTLRWYLATGEIYDPGI 72
QY 316 -HIVQAGNLHSL-----TIAEAFEDTGRYSCFASNIYGTDSAEIYIEGVSSDSEGD 369
DB 73 RHVPNGTLQIFPPPPSSFTLHDNT--YYTAEENSG-KIRSQDVHIAKAV----- 121
QY 370 PNKEEMNRIOKNEV-----SSPPTTSVAVIPPAVQ-----AQH 403
DB 122 -----LREPYTVRVEDQKTMGRGNVAFVKLIIPSSVEAYITVVSWEKDTVSLVSGSRF 173
QY 404 LVAQPRVATQQOCS-----PNTYLOGLDGKPLIAAPVTMKLN 443
DB 174 LITSTGALYIKDVQNEGLNYRCITRHYTGETRQSARLFVSDPANSAPSLDGFH 233
QY 444 LSASEGOLVVFECRKGAPSPKVEWYREGTLIEDSPFRILQKKPRMAPEEICTLVIA 503
DB 234 RKWAGQVVELPCKALGHPEPDYRWLKNMPLSLGRFQ-----KTVTGLLIE 281
QY 504 EVFAEDSGCFTASNKYGTVSSIAQLHVRGNEGLNSLHNSANSTNLAAIEPQPSPP 563
DB 282 NIRPSDSGSYVCEVSNRYGTAKVIGLYVK-----KTVTGLLIE 311
QY 564 HSEPPSVEQPPKPLEGLVYNHNEPRSSRIGLRVHNLDEDDKGSSEASAGVVTTRQT 623
DB 312 -----OPLKATI-----SPRK-----VKSSVGSQVSLSCSTGTGTD- 342
QY 624 RPSXQERF---NGQATKTPSPVKEPPVLAAPKLDSTQLQOLHNOVLLEHQLQNP 680
DB 343 -----QELSWRNGE-----ILNPGK--NVRITGINHENLIMDHVWKS 379
QY 681 PPS-----SPKEFFPMXTVLNSNAPPVATTSKQVKAPSSQTSFLARPKYFPST 730
DB 380 GGAYQCFVRKDKLSAQD--YVQVVELEDGTPKIISAFSEKVVSPAEPVSLMCNVK---GT 433
QY 731 NTAATVAPSSPVFTLSSPTQRTQTVKRESLLVSHPSVQTKS--PGGLSIQNEPLPPG 788
DB 434 PLPTITWLDODDPT--LKGGSHRISQMITSEGNVYVSLTSSSQVRDGV----- 481
QY 789 PTEPTPPPTFSIPSGNQ-----FOPRCVSPVPSPTSRIONVAFSLSSVLP--SLPAIP 841
DB 482 -----YRCTANNSAGVLYQARI-----NVRGP-----ASIRPMKNITAI 517
QY 842 PTNAXMLPRSAPMSQGLAKKNTKSPQVNDNDNIRETKNAVIRDLGKKITFSVRPNQ 901
DB 518 GRDPTIHC-R-VIGYPYYSIKWYKNSLLPFNHRQVAFENNGTLK-----LSDV---QK 566
QY 902 EYKISSFEORLMNEIFRLERTPVDESD--DEIQHDEIPTGKCI-----APIFDKRLKH 953
DB 567 E-----VDEGEYTCNVLPQSLSTSQSQVHVTVKVPPIQPFEEF 604
QY 954 FRVTEGSPVFTCKIVGIPVP-KVYWFKDGKQISKRNECHKMRREG-----DGPCS 1003
DB 605 PRFSIGQVFTPCVVVSGDLPTITWQKDRPI-----PGSLGVITIDNIDFTSS 653
QY 1004 LHIESTSDDDGNYTITMAANPQGRISCSGHLMVQSLPIRSRLTSAGQSHRGRSRVQERDK 1063

Db 654 LRISNLSMHNGNYTCIARNEAAVEHQSLIVR-VP----- 689
Qy 1064 EPLQERFRPHFLOAPGDWAHEGRCLRCDCKVSLPPELTWLLN---GQVLPDASHK 1120
Db 690 -----PKFVQPRQDGIYKAVILNCSAEGYVPTIWKFSKAGVQFPQIALN 740
Qy 1121 MLVRETVGHSLLIDPLTQDAGTYKCIATNKTQGN-SFSLSLSVAKVKKAPVILEKLQ 1179
Db 741 GRIQVLSNGSLLIKHVEEDSGYLVCKVSDVDGADVSKSMYLTV-----KIPAMITSYP 794
Qy 1180 NCGV-PEGHPVRLECRVGMPPPVYWKKNETIPCTRERISMHDGTTG---YACLLIQP 1235
Db 795 NTLTATQCKKEMSCHTARKEPPIVWRKEDRIINPEMAYLVSTKEVGEVISTLQLTP 854
Qy 1236 AKKSDAGWYTLISAKN-----EAGIVSCTARLDIYAQWHHQPPIPMVSRGSGRYSLT-SK 1290
Db 855 TVREDSGFFSCHAINSNGEDRGIIQLTVQ-----EPDPPEIEIKDVKARTILRWTM 907
Qy 1291 GLDIFSAFSSMESTWYSCSSRS 1313
Db 908 GFDGNSPITGYD-----IECKNKS 926

RESULT 10
DCC_MOUSE STANDARD; PRT; 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor suppressor protein DCC precursor.
GN DCC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RC STRAIN-BALB/c; TISSUE-Brain;
RX MEDLINE=96112625; PubMed=8570174;
RA Cooper H.M., Ames P., Britto J., Gad J., Wilks A.F.;
RT "Cloning of the mouse homologue of the deleted in colorectal cancer
RT gene (mDCC) and its expression in the developing mouse embryo.";
RN Oncogene 11:2243-2254(1995).
RN [2]
RP REVISIONS
RC STRAIN-BALB/c; TISSUE-Brain;
RA Cooper H.M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: A (shown here) and B: are
CC produced by alternative initiation. A third isoform: C is
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC ISOFORM C IS EXPRESSED ONLY IN THE EMBRYO.
CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC AND REMAIN AT THIS LEVEL IN THE ADULT.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X85788; CAA59786.1; -.
DR HSP: P56276; I7LK.
DR MGD; MGI:94869; Dcc.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_like.
DR Pfam; PF00041; fn3; 6.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00410; IG_Like; 2.
DR SMART; SM00408; IG2; 3.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 1447
FT CHAIN 85 1447
FT INIT_MET 85 85
FT DOMAIN 26 1097
FT TRANSMEM 1098 1122
FT DOMAIN 1123 1447
FT DOMAIN 154 124
FT DOMAIN 154 219
FT DOMAIN 254 317
FT DOMAIN 345 407
FT DOMAIN 426 522
FT DOMAIN 525 618
FT DOMAIN 619 716
FT DOMAIN 722 816
FT DOMAIN 840 940
FT DOMAIN 941 1042
FT DISULFID 61 117
FT DISULFID 161 212
FT DISULFID 261 310
FT DISULFID 352 400
FT CARBOHYD 60 60
FT CARBOHYD 94 94
FT CARBOHYD 299 299
FT CARBOHYD 318 318
FT CARBOHYD 478 478
FT CARBOHYD 628 628
FT CARBOHYD 702 702
FT VARSPLIC 819 838
SQ SEQUENCE 1447 AA; 158298 MW; 0D1F107C22D5B9F CRC64;

Query Match 4.5%; Score 307; DB 1; Length 1447;
Best Local Similarity 28.3%; Pred. No. 2.4e-07;
Matches 107; Conservative 46; Mismatches 129; Indels 96; Gaps 15;

Qy 959 GSPVTFCKIVGIPVYVFKDGKQISKNEHCKMRREGDG-----TCSLHTESTSD 1012
Db 154 GDTVLLKCEVIGEMPMTIHWKNOQDLN-----LPGDSRVVVLPSGALQISRLQPG 205
Qy 1013 DDGNYTMAANPQGRSCSGHLMVLSQLPIRS-----RLTSAGOSHRSRVSQERDREPL 1066
Db 206 DSGVYRCSARNPAS-----IRGNEAEVRLSDPGLAR----- 238
Qy 1067 QEREFRRPHFLOAPGDWAHEGRCLRCDCKVSLPPELTWLLNQPVLDPASHKMLVRET 1126
Db 239 -----QLYFLQRPNSVIAIEGKDAVLECCVGYPPSFTW-LRGEVITQLRSKYS-L 290
Qy 1127 GVHSLLLDPLTQDAGTYKCIATNKTQGN-SFSLSLSVAKVKKAPVILEKLQNCGVPEG 1186
Db 291 GGSNLLISNVTDDSGTYTCTVYTKENISASAEITLV-----PPWFLNHPNSLYAES 345
Qy 1187 HPVLECRVIGMPPPVYWKKNDE-TIPCTRERISMHDGTTGAYCALIQPAKKSADAGWT 1245
Db 346 MDIEFECVAGSKGPVPTVNMWKNQDVIP-----SDYFQIVGGSNLRILGVKRSDEGFYQ 399
Qy 1246 LSAKNEAGIVSCTARLDIYAQWHHQP-----PPMSVRPS-----GSRYGSLT--- 1288

Db 400 CVAENEAGNAQSAQL-----IVPKPAIPSSILPSAPRDVLPVLVSSREVRLSWRP 451
QY 1289 ---SKG-LDIFSASFSSME 1302
Db 452 PABAKGNIQTFTVFFSRE 469

RESULT 11
MPSE_CHKID
ID MPSE_CHKID STANDARD; PRT; 1450 AA.
AC Q02173;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE M-protein, striated muscle.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic pectoralis muscle;
RX MEDLINE=93015907; PubMed=1400348;
RA Noguchi J., Yanagisawa M., Imamura M., Kasuya Y., Sakurai T.,
RA Tanaka T., Masaki T.;
RT "Complete primary structure and tissue expression of chicken
RT pectoralis M-protein.";
RL J. Biol. Chem. 267:20302-20310(1992).
CC -!- FUNCTION: IS A STRUCTURAL CONSTITUENT OF MYOFIBRILLAR M-BAND IN
CC STRIATED MUSCLE.
CC -!- TISSUE SPECIFICITY: IS EXPRESSED IN PECTORALIS AND CARDIAC MUSCLE.
CC -!- DEVELOPMENTAL STAGE: CAN BE DETECTED BY DAY 10-13 IN OVO. THE
CC CONTENT IS GRADUALLY INCREASED THROUGHOUT THE OVO DEVELOPMENT AND
CC REACHED ITS PEAK AFTER HATCHING.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D11474; BAA02033.1; -.
DR PIR; A44027; A44027.
DR HSP; P56276; ITLK.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FnIII_repeat.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; ig; 5.
DR PRINTS; PRO0014; FNTYPEIII.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00410; IG_like; 3.
DR SMART; SM00408; IGC2; 1.
KW Muscle protein; Repeat; Immunoglobulin domain.
FT DOMAIN 142 238 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 266 364 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 368 464 FIBRONECTIN TYPE-III 1.
FT DOMAIN 496 592 FIBRONECTIN TYPE-III 2.
FT DOMAIN 597 693 FIBRONECTIN TYPE-III 3.
FT DOMAIN 696 794 FIBRONECTIN TYPE-III 4.
FT DOMAIN 797 898 FIBRONECTIN TYPE-III 5.
FT DOMAIN 899 990 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 1002 1095 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 1126 1205 IG-LIKE C2-TYPE DOMAIN 5.

FT DOMAIN 1225 1312 IG-LIKE C2-TYPE DOMAIN 6.
FT DOMAIN 1333 1427 IG-LIKE C2-TYPE DOMAIN 7.
SQ SEQUENCE 1450 AA; 163400 MW; E313C7D39FE6CAC7 CRC64;
Query Match 4.4%; Score 301.5; DB 1; Length 1450;
Best Local Similarity 18.6%; Pred. No. 4.3e-07;
Matches 292; Conservative 183; Mismatches 516; Indels 577; Gaps 65;
QY 143 ETOSKKVFLNKAADFTLELSSLFKSH---SSKRIRPRACKNHHKSLKESQNVKQENSSSF 199
Db 31 EYARRKAASQAAHY--ESTGLGKTTCRLCARRARSLA---HEAMQESRRKT-HEOKSHA 84
QY 200 SDSLERR-----ERSSVPIPIPADTRDNEY-----NHALEQOE---AKRREAQAAASAAG 247
Db 85 SD--EKRIKFASELSSL-----EREIHMARHAREQLDRLAIQRYVEENWALERH- 132
QY 248 GDTTPGSSPSLYEEPLGQPPRTOKLRSREVEPTGRVQLDICIVGIPPPQVWVCEGK 307
Db 133 -----VVEEKISRAPELLVRLSHTVMEKMSVRLCTVCGFPSPVQWY-KNE 179
QY 308 EL-----ENSPDIHIQVAGNLHSLTAAFEEDTGRYSCFASNIYGTDSAE 355
Db 180 ELITPASDPKAKYVENKYGIVHV-----LHINRADFDSATYSAVATNIHQOASTNCA 231
QY 356 IYIEGVSSSDSEGDPNKEEMNRIQKPNESVSPPTTSAVIPPAVPAQOHLVAQPRVATIQ 415
Db 232 VVYRRPRESE-----EHPAGIMP 250
QY 416 CQSTNYVLOGLDGKPIIAAPVFTKMLQNL-----ASEQLVFCRVKVGAPF-----PKVE 467
Db 251 FHLPLSY-----DVCFTFDVQVLEKFGVTEGETITLKCSSLVLTPELRLRPRAE 303
QY 468 WYREGTLIEDSPDFRILQKKPRSMAPPEICTLVIAEFAEDSCGCTCTASNYGTVSSI 527
Db 304 WYRDDVLIKDS-----KWTXLYFEGEQAALSFTHLNKDDDELGYLTRMVTKGGVNECS 355
QY 528 AOLHVRGNEDLSNNG-----SLHSANSTNLAIAIPQSPHSEPPSVEQPPKPKLEG 580
Db 356 AFLFVRDADALIAAGAPGAPMDVKCHDANRDYVIVTWKP-----DNTTSQNP---VIG 404
QY 581 VLVNHNPRSSRIGLR--VHFNLPEDDKGSEASSEAGVVTTRQTRPDSQER---FNGQ 635
Db 405 YFVDKCE-----VGLNWMYQCN-----DAPVKICKYPTVTLGEGRSYIFRVR 446
QY 636 ATKTPSEFVKPPPPVLAKPKLDSTOLOQLH---NOVLLEQHLQNP-----PPSSPK 686
Db 447 AVNSAGISRSRSEPVAAALDPVDLERTQTVHVDGGRKIVISKDDLEGGDIQPGPTNVH 506
QY 687 EFPFXMTVLNSNAPPVATTSXKQVKAPSSQTSFLARPKYFFPSTNTTAATVAPSSS--PV 744
Db 507 ASEISKTYVVLSDWDPVPVGRGRL-----TYFIEKSMVSGSGWQRYNAQVAKSPRYAV 560
QY 745 FTLSSSTPQTITQTVSKESLLVSHPSVQTKSPGGLSIQNEPLPGPGTPTPTPPFT----- 798
Db 561 FDLAEGKPYFVYVLSANKHGIDSPSEITEPIQODIVVWPSPAPGRVAVTRNTKTSVVVQW 620
QY 799 -----FSPISGNQFQP-----RCVSPIPV 817
Db 621 DKPKHEENLYGYIDYSVSGSNQWEPANHKPINYRNVFVHGLETGEQYIFRVKAVNAVGF 680
QY 818 SPTSRIQNPAFLSSV-LPSLP-AIPPTN---AMXLPRSAPSMPSQGLAKNKTSP--- 868
Db 681 SENSESEAIKVAQALTCPSYPHGICITLLNCDGHSMTLQWKAP-----KYSGSGPILG 732
QY 869 -----QPVNDNIRETKNAVI-----RDLGK-----KITFSDV-----RPN--QOEYK 904
Db 733 YYIDKREANKHNWHEVNSSVISRTIYTVEDLTEDAFYEKIAAANVIGIGHPSPDSEHK 792
QY 905 ISSFEQ-----RLMNEIEFLERTPVDE-----SDDEIOHDE 936
Db 793 CKAWTMPGPGAYDLTVCEVRNTSLVLLWKAPVYEGKSPITGYLVLDYKEVDTEDMITANE 852
QY 937 IPTGKCIAPFDKRLKH---FRV-----TEGSPVT----- 963

```
Db 853 KPTSHRYFKVTLHQGHTYVFKVRAVNDAGVGKSEISEPVEAFSPGCTKEIFSGVDEEG 912
Qy 964 -----FTCKIVGIPVKYVFKDGCKQISKRNECHCKMRREGDGTCSLHTESTTSDDDGNYT 1018
Db 913 NIYLGFECK-EATDASHFLWCKGSYEEL-EDSDKFKIETFKGDHS-KLYFKHPDKSDLGTYC 969
Qy 1019 IMAANPOG-----RISCSCHLMVQSLPRLSRL----- 1045
Db 970 ISVSDTDGCVSSFFVLDEEELRLMTLSNEIKNPITPLKSELAYELVDKGEVRFWIAESL 1029
Qy 1046 -----TSAGQSHR-----GRSRVQERD----- 1062
Db 1030 SPNSTYRFVINDKEVNGDRHKISCDHSHNGIEMVMDKFTIDNEGTYTVQIQDGKAKNQ 1089
Qy 1063 -----KEPQERFR-----PHELOAPGDMAHEGLRCL--LDCKVSGGLPPP 1102
Db 1090 SLVLIGDAFKAILAESELOREFLRKQGFHSEFLYWEVTEE---CEVLLACKTANTYKKE 1146
Qy 1103 EL-TWLLNGQVPLPDASHKMLVRETGVHSLLDPLTQDAGTYKCIATNKTGQNSFLEL 1161
Db 1147 TVFKWYRNGSGIDVEAPDL---QKGECHLTVPKLSRKDEGVYKATLSDDRGHDVSTLEL 1203
Qy 1162 S-----VVA-----KEVK----- 1169
Db 1204 SGKVNDIILALSRVSGKTASPLKILCTEIGIRLQCFLKYYNEEMKVWTSRESKISSGE 1263
Qy 1170 ----- 1171
Db 1264 KMKTGGEDVAVLQITEPTEKDKGNKYTFEIPSDKESKRTLDLGSQAFDDALTEFFQLKA 1323
Qy 1172 PVILEKLQN---CGVP-----EGHPVRLCEVRVIGMPPPVFVYKWKDNETIPCTRE-RISM 1221
Db 1324 AAFAEKNRGKVIYGLPDVVTIMDKTCLNLTCTVFGNPDPEVWFVFNKDKALENHELYLVS 1383
Qy 1222 HQDTTGTACLLIQAPKSDAGWYPLSAKNAGIVSCFARLDIYAQWHH-----QTPPP 1274
Db 1384 EQGK--YASLTIKGTSEDGSKYIYVKNKYGETVDVTVSVRHGEKIPEVNOGLAKP 1441
Qy 1275 MSVRPSS 1282
Db 1442 RLIPSSS 1449

RESULT 12
DCC_HUMAN
ID DCC_HUMAN STANDARD; PRT; 1447 AA.
AC P43146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor suppressor protein DCC precursor (Colorectal cancer suppressor),
GN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95011532; PubMed-7926722;
RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA Vogelstein B.;
RT "The DCC gene product in cellular differentiation and colorectal
RT tumorigenesis."
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RP SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE-90100559; PubMed-2294591;
RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA Vogelstein B.;
RT "Identification of a chromosome 18q gene that is altered in
RT colorectal cancers."
Science 247:49-56(1990).
[3]
RN RN
RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBELD EXONS).
RX MEDLINE-91121517; PubMed-1991322;
RA Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA Oliner J.D., Kinzler K.W., Vogelstein B.;
RT "Scrambled exons."
RL Cell 64:607-613(1991).
[4]
RN RN
RP GENE STRUCTURE AND VARIANTS CARCINOMA HIS-1375.
RX MEDLINE-94245241; PubMed-8186295;
RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT "The DCC gene: structural analysis and mutations in colorectal
RT carcinomas."
RL Genomics 19:525-531(1994).
[5]
RN RN
RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX MEDLINE-94243823; PubMed-8187090;
RA Miyake S., Nagai K., Yoshino K., Ota M., Endo M., Yuasa Y.;
RT "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT human esophageal squamous cell carcinomas and their relation to
RT metastasis."
RL Cancer Res. 54:3007-3010(1994).
[6]
RN RN
RP FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC CC
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC CC
CC -1- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC CC
CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC CC
CC -1- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC CC
CC DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC CC
CC EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC CC
CC POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC CC
CC METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC CC
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC CC
CC SUBFAMILY.
CC CC
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC CC
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
EMBL; X76132; CAA53735.1; -
DR EMBL; M32292; AAA52174.1; -
DR EMBL; M32286; AAA52174.1; -
DR EMBL; M32288; AAA52175.1; ALT_SEQ.
DR EMBL; M32290; AAA52176.1; -
DR EMBL; M63696; AAA52177.1; -
DR EMBL; M63700; AAA52178.1; -
DR EMBL; M63702; AAA52179.1; -
DR EMBL; M63718; AAA52180.1; -
DR EMBL; M63698; AAA52181.1; -
DR PIR; A54100; A54100.
DR PIR; A40098; A40098.
DR PIR; A38442; A38442.
DR HSP; P56276; IYTK.
DR Genew; HGNC:2701; DCC.
MIM; 120470; -
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_C2.
DR InterPro; IPR003600; Ig_Like.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00047; ig; 4.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00410; IG_Like; 2.
DR SMART; SM00408; IGC2; 3.
```


entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; AF230073; AAC00757.1; -
CC DR EMBL; AL109804; CAC17543.1; -
CC DR EMBL; AL109804; CAC17542.1; -
CC DR EMBL; AK024462; BAB15752.1; -
CC DR EMBL; AK024459; BAB15749.1; -
CC DR EMBL; AK024479; BAB15769.1; -
CC DR HSSP; Q62230; 10FO.
CC DR Genew; HGNC:11127; SN.
CC DR MIM; 600751; -
CC DR InterPro; IPR003599; Ig.
CC DR InterPro; IPR003006; Ig_MHC.
CC DR InterPro; IPR003598; Ig_c2.
CC DR InterPro; IPR003600; Ig_1like.
CC DR Pfam; PF00047; Ig; 14.
CC DR SMART; SM00409; Ig; 17.
CC DR SMART; SM00410; Ig_1like; 1.
CC DR SMART; SM00408; IGG2; 14.
CC KW Signal; Glycoprotein; Cell adhesion; Immunoglobulin domain; Repeat;
CC KW Transmembrane; Alternative splicing; Antigen.
CC FT SIGNAL 1 19 BY SIMILARITY.
CC FT CHAIN 20 1709 SIALOADHESIN.
CC FT DOMAIN 20 1641 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 1642 1662 POTENTIAL.
CC FT DOMAIN 1663 1709 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 20 136 IG-LIKE V-TYPE DOMAIN.
CC FT DOMAIN 153 224 IG-LIKE C2-TYPE DOMAIN 1.
CC FT DOMAIN 255 312 IG-LIKE C2-TYPE DOMAIN 2.
CC FT DOMAIN 339 397 IG-LIKE C2-TYPE DOMAIN 3.
CC FT DOMAIN 426 498 IG-LIKE C2-TYPE DOMAIN 4.
CC FT DOMAIN 524 582 IG-LIKE C2-TYPE DOMAIN 5.
CC FT DOMAIN 617 696 IG-LIKE C2-TYPE DOMAIN 6.
CC FT DOMAIN 722 781 IG-LIKE C2-TYPE DOMAIN 7.
CC FT DOMAIN 810 883 IG-LIKE C2-TYPE DOMAIN 8.
CC FT DOMAIN 909 967 IG-LIKE C2-TYPE DOMAIN 9.
CC FT DOMAIN 998 1074 IG-LIKE C2-TYPE DOMAIN 10.
CC FT DOMAIN 1100 1156 IG-LIKE C2-TYPE DOMAIN 11.
CC FT DOMAIN 1186 1248 IG-LIKE C2-TYPE DOMAIN 12.
CC FT DOMAIN 1274 1331 IG-LIKE C2-TYPE DOMAIN 13.
CC FT DOMAIN 1360 1432 IG-LIKE C2-TYPE DOMAIN 14.
CC FT DOMAIN 1458 1518 IG-LIKE C2-TYPE DOMAIN 15.
CC FT DOMAIN 1547 1620 IG-LIKE C2-TYPE DOMAIN 16.
CC FT DISULFID 41 98 BY SIMILARITY.
CC FT DISULFID 160 217 BY SIMILARITY.
CC FT DISULFID 262 305 BY SIMILARITY.
CC FT DISULFID 346 390 BY SIMILARITY.
CC FT DISULFID 433 491 BY SIMILARITY.
CC FT DISULFID 531 575 BY SIMILARITY.
CC FT DISULFID 624 689 BY SIMILARITY.
CC FT DISULFID 729 774 BY SIMILARITY.
CC FT DISULFID 817 876 BY SIMILARITY.
CC FT DISULFID 916 960 BY SIMILARITY.
CC FT DISULFID 1005 1067 BY SIMILARITY.
CC FT DISULFID 1107 1149 BY SIMILARITY.
CC FT DISULFID 1193 1241 BY SIMILARITY.
CC FT DISULFID 1281 1324 BY SIMILARITY.
CC FT DISULFID 1367 1425 BY SIMILARITY.
CC FT DISULFID 1465 1511 BY SIMILARITY.
CC FT DISULFID 1554 1613 BY SIMILARITY.
CC FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 697 697 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 730 730 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 886 886 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1138 1138 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 1251 1251 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1462 1462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1476 1476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1632 1709 ALHRLHOFQOLLWGLLVGLLLGLLGLGACYTWRRRVCK
FT QSMGENSEMAFOKETQLIDPDAACTETSCAPLG ->
FT GEGRLHLPCHSAQKFPSS (IN ISOFORM 2).
FT CONFLICT 1349 1349 A -> T (IN REF. 1).
FT CONFLICT 1519 1519 A -> V (IN REF. 3; BAB15749/BAB15769).
SQ SEQUENCE 1709 AA; 182624 MW; 587C7CA0B789A6D CRC64;

Query Match 4.38; Score 296.5; DB 1; Length 1709;
Best Local Similarity 20.3%; Pred. No. 9e-07;
Matches 276; Conservative 143; Mismatches 492; Indels 451; Gaps 57;

QY 107 HSPNLSFEPNFCQDNPRSTSSKESPOBAKRPOY-CSETOSKKVFNKAA---DFIEELS 162
DB 461 HSPRFS-----GTSGPNSLRLEIRDLEETDSGEYKCSATNS---LGNATSTLDFHANAA 511

QY 163 SLFKSHSKRIRPRA-----CKNHKSLESQNKVMQENSSSFSDLSERRRERSVPIPAD 218
DB 512 RLLISPAAEVVEGQAVTLSCRSGLSTPDPARFSWYLNKA-----LLHEGPGSLLLPAASS 567

QY 219 TRDNEVNHAEQOQAKRREAQAASEAAGDTTPG-SSPSLYEEPLGQPPRTQK--L 275
DB 568 T-----DAGSYHCRARDGHSASGSPSPAVLVLYPPRO-PTFTTLLDL 609

QY 276 RSREVPEGTRVOLDICIVVGIPPPQVRWYCEGKELEN-----SPDTHIVQAG 321
DB 610 DAAGAGARRGLLLCRVDSDPPARLQLLHKDRVVATSLPSGGGCGTCGGSPRKNVTKAP 669

QY 322 NLHSLTIAEAFEDTGRYSCFASNIYTDSTSA-----EIVEGVS----- 362
DB 670 NLLRVEIHNPILLEEGLYLCEASNALGNASTSATFNGQATVLAIPASHTLQEGTEANLTC 729

QY 363 --SSDSEGDNPKEEMNKIQKNEV---SSPPTTSVAVTPPAVPOAHLVAQPRVATQOCC 417
DB 730 NVSREAAAGSPANFSWFR-----NGVLWAQGPLETTLTPVARTDAALYAC--RLTTEAGAQ 783

QY 418 SPTNYLQGLDGKPIIAAVFTKMLQNLASBGLVWFECRVKGAQSPKVEWYREGTLIED 477
DB 784 LSTPVLLS-----VLYPDRPKLSALLDMGGMALFICIVDSRPLALLALFGEHLLAT 838

QY 478 S-----PDFRILQKKRSMAPBEICTLVIAEVAEDSGGCTCTASNKYGTVSIAQLHV 532
DB 839 SLGPQVPSHGFRFOAK-----AEANSL-KLEVRELGLGSGSYRCBATVNLGSSNTSLRFQV 893

QY 533 RGNEDLSNGLSHSANTSTNLAATEPOQSPPHSPPEPPVEQPPKPLEGLVNVHNEPRSSS 592
DB 894 RG-----AWQVSPSELQEGQAVV----- 913

QY 593 RIGLRVHFNLPED-----DKGSEASSEAGVYVTTTQTRPDSXQERFNGQATKTP 640
DB 914 -LSCQVHTGVPGETSYRWYRDGQPLQESTSATLRFAAITLTQAGAYHCQAQAPGSATTS- 971

QY 641 EPSPPVKEPPPPVLAKP-----KLDLS---TQLOQLH 667
DB 972 -----LAAPISLHVSYAPRHVTLTTLMDTGPGRGLGLLLCRVDSDPPAQLRLH 1019

QY 668 NOVLEQ--HQLONPSPSPKEFPFXMTVLNSNAPPAVTTXKQVKAQSPQTSFLARPKY 725
DB 1020 GDRLVASTLOGVGGESSPR-----LHVAVAPNTLRLEIGHAMLEDEGV----Y 1065

QY 726 FFPSTNTAATVAPSSSPVFTLSSTPQTIQRTVSKEILLV-----SHPSVOT----- 772
DB 1066 ICEASNTLGO--ASASADFDAQAVNVQVWPGATVREQLVNLCLVWTTTHPAQLTYTWYQ 1123

QY 773 KSPGGLSIQNEPL-----PPGPTTEPPPTTFTSIPSGNQQPRCVSPIP 816
DB 1124 DGOORLDAHSIPLNVTRDATSYRCGVPGRAPLSRPTLDV-----LYAPRN 1174

QY 817 VSPITSRQNPAFLSSVLPSPILPAPP-----TNAMXILPRAPSMPQGLAKKNTKSPQV 871
DB 1175 LRLTYLLESHGGQALVLCTVDSRPPAQLALSHAGRLLASSTAASVPNTLRLEURGPQP- 1233


```
QY 872 NDNIRETKNAVIRDLGKKITFSDVRNQQEYKISSFEQRLMNEIFRLERTPTVDESDD 931
Db 1234 -----RDEG-----FYSCSA-----RSPGLQANTS 1253
QY 932 IOHDEIPTGKCIAPIDFKRLKHFVTEGSPVTTCTKIVGIPVKVY-WFKDQKQISKRNE 990
Db 1254 LEURLEGVRILAP-----EAAVPEGAPITVTCADPAHAHPTLXTWYHNGRWL----- 1301
QY 991 HCKMRREGDCTSLHIBSTSDGGNTVMAANPOGRISCGHLMVQSLFIRSLTSAGQ 1050
Db 1302 -----QEGP--AASLSFVATRAHAGAYSCQAQAQG----- 1331
QY 1051 SHRCRSVQERDKPELQERFRPHFLQ---APGDMVAHEGRLCR-----LDCKVSLGPP 1101
Db 1332 -----TRSSRAALQVLPQDAVLSSFRDSRARSMAVICTVDSEPP 1374
QY 1102 PELTWLNGOPVLDPASHKMLVRETGVHSLIDP-----LTQRD--AG--TYKC 1146
Db 1375 AELALSHDG-----KVLATSSGVHSLASGTGHVQVARNALRLQVQDVPAGDDTYVC 1425
QY 1147 IATNKTGNSFSLSLVAK-EVKAPVILEKLQNCVPEGHPVRLRCRVIGMPPV--- 1202
Db 1426 TAQNLG-----SISTIGRLQVEGARVAE--PGLDVPAGALNSCRLLGPGPGVNS 1477
QY 1203 ---FYW---KKNETIPTCTRERISMHDQTTGYACLLIQPAKKSADAGWYTLKSAKEAGIVS 1256
Db 1478 TEAFWENDRLHAEPVPTLA-----FTHVARAQAGMYH 1510
QY 1257 CTARLDIYAQWHQI-----PPP-----MSVRPSGRSRYGSL 1287
Db 1511 CLAEPLTGAASAPVMLRVLYPKPTPTMMVFVEPEGGLRGIL 1552

RESULT 14
MYM1_MOUSE STANDARD; PRT; 1666 AA.
AC Q62234;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myomesin 1 (Skelemin).
GN MYOM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10J; TISSUE=Muscle;
RX MEDLINE=94012766; PubMed=8408035;
RA Price M.G., Gomer R.H.;
RT "Skelemin, a cytoskeletal M-disc periphery protein, contains motifs
RT of adhesion/recognition and intermediate filament proteins.";
RL J. Biol. Chem. 268:21800-21810(1993).
RN [2]
RP INTERACTION WITH BETA-INTEGRIN.
RX MEDLINE=99074282; PubMed=9857037;
RA Reddy K.B., Gascard P., Price M.G., Negrescu E.V., Fox J.E.B.;
RT "Identification of an interaction between the M-band protein skelemin
RT with beta1-integrin subunits. Colocalization of a skelemin-like protein
RL J. Biol. Chem. 273:35039-35047(1998).
CC -1- FUNCTION: MAY LINK THE INTERMEDIATE FILAMENT CYTOSKELETON TO THE
CC M-DISC OF THE MYOFIBRILS IN STRIATED MUSCLE. MAY ALSO CONTACT
CC MYOSIN FILAMENTS. ALSO BINDS BETA-INTEGRINS.
CC -1- TISSUE SPECIFICITY: MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z22866; CAAB0488.1; -.
CC DR SWISS-2DPAGE; Q62234; MOUSE.
CC MGD; MGI:1341430; Myom1.
CC DR InterPro; IPR003961; FN_III.
CC DR InterPro; IPR003962; FNIII_repeat.
CC DR InterPro; IPR003006; Ig_MHC.
CC DR InterPro; IPR003598; Ig_C2.
CC DR InterPro; IPR003600; Ig_Like.
CC DR Pfam; PF00041; fn3; 5.
CC DR Pfam; PF00047; ig; 5.
CC DR PRINTS; PRO0014; FNTYPEIII.
CC DR SMART; SM00060; FN3; 4.
CC DR SMART; SM00410; IG_Like; 4.
CC DR SMART; SM00408; IGC2; 1.
CC KW Immunoglobulin domain; Muscle protein; Thick filament; Repeat.
CC FT DOMAIN 271 334 IG-LIKE C2-TYPE DOMAIN 1.
CC FT DOMAIN 398 463 IG-LIKE C2-TYPE DOMAIN 2.
CC FT DOMAIN 489 575 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 617 703 FIBRONECTIN TYPE-III 2.
CC FT DOMAIN 718 802 FIBRONECTIN TYPE-III 3.
CC FT DOMAIN 915 1004 FIBRONECTIN TYPE-III 4.
CC FT DOMAIN 1020 1108 FIBRONECTIN TYPE-III 5.
CC FT DOMAIN 1136 1193 IG-LIKE C2-TYPE DOMAIN 3.
CC FT DOMAIN 1352 1409 IG-LIKE C2-TYPE DOMAIN 4.
CC FT DOMAIN 1568 1629 IG-LIKE C2-TYPE DOMAIN 5.
CC SQ SEQUENCE 1666 AA; 185330 MW; 6B4393AA169B2D5B CRC64;

Query Match 4.1%; Score 282.5; DB 1; Length 1666;
Best Local Similarity 19.4%; Pred. No. 4e-06;
Matches 294; Conservative 166; Mismatches 502; Indels 557; Gaps 64;

QY 96 ARKRLSPDQMKHNSPLSFEPNFCODNPSPTSSKESQO---EAKRPQYCSFQO----- 146
Db 94 SRKTASAYDYGSHGLTDSLLIED-----YSSKLSQTKRAKVLSGEETGSLPGNYLV 148
QY 147 -----KKVFLNKAADFTIEELSSLFKSHSKRIRPRACK-NHKSKLESQNKVQBNSSFS 200
Db 149 PIYSGRQVHISGIRDSEE-----RIKEAAAYIAQKTLLEAEEAIAASKQSTAS 197
QY 201 DLSERRRSPVPIPADTRDNEVNHAEQEAQKRRAEQAASAAAGDTPPGSSPSLS 260
Db 198 KQSATSKRTT-----STLQREETFEEKSRNIAIREKAEELSLKKTLEET-----QTY 244
QY 261 Y----EPLGQPPRFTOKLRSEVPECTRYQLDCIVVGIPPOVVRWYCEGK--ELENSPD 314
Db 245 HGKLNEDHLLHAPEFIKPRSHTWKENVKLHCSVAGWPEPLTWYKNQVPIVNHANPG 304
QY 315 IHIVQAG-NLHSLTIAPAEEDTGRYSCFASNIYGTDSIAEYIEGVSSSDSGDPNKE 373
Db 305 KYIIESRYGMHTLEISKDFEDTAQYRASAMNVQGLSAYASVVVKRY-----KG 354
QY 374 EMRIQRPNEVSSPPTTSVAVIPPAVPOAHLVAQPRVATIQOCOSPTNYLQGLDCKPIIA 433
Db 355 ELDESLLRGVSM-PLSFAVTP-----YGYASKVEIH 385
QY 434 APVFTKMLQNLASASEGOLVVFECRVKGP-----SPKVEWYRECTLIEDSPDFILQKP 488
Db 386 ---FDDKFDVSGREGETMSLGCRCVITPELKHQFQVQWRNGAPV-----SP 431
QY 489 RSMAPPE---EICTLVIAEYFAEDSGCTCTA--SNKYGTVSSIAQLHVRGNEDLSNNGS 543
Db 432 SKWQPHWSGDRATLTFSLHNKKEDEGLYTIIVRMGEYEQYS--AYVVRDAD----- 482
QY 544 LHSANSTNLAIEPQPSPPHSEPPSVEQPPKPLEGLVNVNHNPRSSRSLGLRVHNL 603
Db 483 -----AEIEGAPAAP-----LDVVSLDANKDY-----IISWKQP 512
QY 604 EDDKGS-----EASSEAGV- 617
```


GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 30, 2002, 19:07:40 : Search time 397 seconds
(without alignments)
7487.753 Million cell updates/sec

Title: US-09-818-990b-2

Perfect score: 6890

Sequence: 1 MQDDSEASTSISQLRESY.....MESTWVSCSRSSVVEDEL 1320

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame_p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09818990/runat_26112002.093404.23209/app_query.fasta_1.1479
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1.1.277 @runat_26112002.093404.23209 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2560	37.2	2127	22	AAH13889	Human cDNA sequenc
2	2164	31.4	1228	23	AAH13889	DNA encoding novel
3	1510.5	21.9	1823	21	AAC98221	Human colon cancer
4	1421.5	20.6	4125	21	AAC6971	Human ORFX ORF2526
5	1394	20.2	2059	22	ABA06552	Human cDNA SEQ ID
6	1380	20.0	2232	22	ABA06771	Human immunoglobul
7	1133.5	16.5	702	22	ABA06771	Human cDNA SEQ ID
8	1133.5	16.5	702	22	AAH28841	Human immunoglobul
9	922.5	13.4	855	22	AAH04212	Human cDNA clone (
10	810.5	11.8	81940	24	ABK64829	Human benign prost
11	806.5	11.7	81940	22	AAH03390	Human titin (conne
12	763	11.1	790	20	AAH00026	Prostate cancer as
13	760	11.0	2244	21	AAH61262	cDNA encoding a hu
C 14	668.5	9.7	747	20	AAH40027	Prostate cancer as
15	656	9.5	392	21	AAH75084	Human ORFX ORF639
16	581.5	8.4	15272	23	ABL03697	Drosophila melanog
17	557.5	8.1	5926	24	ABN96874	Gene #372 used to
18	557.5	8.1	5926	24	ABL61910	Colon adenocarcino
19	557.5	8.1	5926	24	ABL62450	Colon adenocarcino
20	557.5	8.1	5926	24	ABL65212	Lung cancer relate
21	557.5	8.1	5926	24	ABL69436	Prostate cancer re
22	557.5	8.1	5926	24	ABK35571	Gene MYLK differen
C 23	552.5	8.0	26097	23	ABL03696	Drosophila melanog
24	546.5	7.9	16908	24	ABQ86156	Novel human gene.
25	523.5	7.6	1768	24	AAH59901	Polynucleotide enc
26	496	7.2	20448	23	ABL19989	Drosophila melanog
27	487.5	7.1	4611	23	ABV24911	Human prostate exp
28	485.5	7.0	4611	23	ABV23407	Human prostate exp
29	485.5	7.0	4611	23	ABV29265	Human prostate exp
30	481.5	7.0	12777	23	ABL28641	Drosophila melanog
31	481.5	7.0	24789	23	ABL28640	Drosophila melanog
32	464.5	6.7	16421	23	AAH86623	DNA encoding novel
33	462.5	6.7	3910	23	AAH68120	DNA encoding novel
34	455.5	6.6	2508	22	AAH57454	Human liver cell s
C 35	441.5	6.4	24971	23	ABL19988	Drosophila melanog
36	432.5	6.3	3192	21	AAF16067	Human prostate can
37	431	6.3	14327	22	AAF54724	Nucleotide sequenc
38	429.5	6.2	14586	23	AAH87452	DNA encoding novel
39	428	6.2	3369	23	ABL19623	Drosophila melanog
40	426	6.2	23546	23	ABL02655	Drosophila melanog
41	416.5	6.0	8882	21	AAZ36321	Mechanical stress
42	416.5	6.0	8883	24	AAI72583	Rat OCP cDNA. Rat
43	413.5	6.0	7679	21	AAZ36322	Mechanical stress
44	408.5	5.9	8262	24	AAI72586	Human OCP cDNA. H
C 45	393.5	5.7	6192	23	ABL19622	Drosophila melanog

ALIGNMENTS

RESULT 1

AAH13889
ID AAH13889 standard; cDNA; 2127 BP.

AC AAH13889;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:10895.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 10895; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2127 BP; 638 A; 501 C; 487 G; 501 T; 0 other;

Alignment Scores:
Pred. No.: 5,71e-120 Length: 2127
Score: 2560.00 Matches: 494
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.16% Indels: 0
DB: 22 Gaps: 0

US-09-818-990b-2 (1-1320) x AAH13889 (1-2127)

Qy 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 231 ATGCAAGACGACACATAGAGCTTCTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 290
Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 291 TTAGCTGAACACACATATCGGGGAAACAAATGAGAGAGTTCGAGCGGAGCCCTCTCTCAAC 350
Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeu 60
Db 351 CTTTGCCATTTTCGGCAGCTCTCTCTGGGGCCGCTGAAGAGGCGGAGGCCAAGATGACCTT 410
Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 411 CCAGATCTTTCAGGCTTCTTGAGGCCAAGAGAAATTAGACGAAAGTGTCAATTTGGCAAGA 470

Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Db 471 CTGGCCATCAATTACGACCTTTGGAGAAGCGAGATCAAACTCAAGCTAGAAAAACGACTT 530
Qy 101 SerProaspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Db 531 TCTCCTGATCAGATGAACACTCACCTAATTTAAGTTTGTAGCCTTAACCTTCTGCCAGGAT 590
Qy 121 AsnProArgSerProThrSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Db 591 AACCTCTGAAGTCCACACGCTCTAAAGAAAGCCCCAGAGGAGGCAAAAAGGCCACAGTAT 650
Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Db 651 TGTCTCAAAACCCAGTCCAAAAAAGTATTTTAAATAAGGCTGCCGACTTCATTGAAGAG 710
Qy 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
Db 711 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCGTGCCTGCGCAAAAC 770
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAlaLysSerSerPheSer 200
Db 771 CACAGAGTAAACTGGAAATCTCAAAACAAGATTATGCGAGGAAACACAGCTCAGTTTCTCA 830
Qy 201 AspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
Db 831 GATCTGTCAAGAACAGCAGAGAAAGATCTTCTGTCCCATCCTATCCTCGCGATACCAGG 890
Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
Db 891 GATAATCAAGTGAATACGCGCTTGAACAGCAGAGCAAGAGCGGTGAAGCGGAGCAG 950
Qy 241 AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
Db 951 GCTGCCAGTGGCGGCTGGTGGAGACACTACACCGAGGCTTCCCTTCTATCTCTGTAC 1010
Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
Db 1011 TATGAAGAAGCTCTGGGGCAACCTCCCGGTTCACTCAAAAGTTACCGGAGCAGAGAAGTT 1070
Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGluVal 300
Db 1071 CCAGAAGAACTCTGAGTACAGTTGGATTGATAGTGTAGGAATTCACCACTTCAAGTA 1130
Qy 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
Db 1131 AGGTGGTACTGTGAAGCAAGGAGCTTGAAATTCCTCCAGATATTACATCGTCCAGCA 1190
Qy 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
Db 1191 GGAAATCTGCACCTCACCTGACCATTTGCGGAAGCCCTTTGAAGAGGACACAGGACGCTATTC 1250
Qy 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
Db 1251 TGTCTTCTTCAACATCTATGGGACAGATTCGACTTCTGCTGAGATTATATAGAAGG 1310
Qy 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
Db 1311 GTTTCTTCTTCTGACTCAGAAGGCGACCCCTAACAAGGAAGAGATGAATCGAATCCAGAAG 1370
Qy 381 ProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGln 400
Db 1371 CCAATGAGGTGTCATCTCCTCCACCTACCTCTGCAGTCAATTCCTCCAGCAGTAGTACCCAA 1430
Qy 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
Db 1431 GCCCAGCATTTGGTGGCCCAACCTCGTGTGCAACCATCCAGCAGTGTGAGAGCCCCACC 1490
Qy 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMet 440
Db 1491 AATTACTTGAGGATGGATGGAAACCTTATCATTCGACGCTCTGTGTGTACAAAGATG 1550
Qy 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460

|||||
Db 1551 CTACAAATTTGTCAGCTTCTGAGGGTCAGCTGGTTGCTCTTTCAATGCAGAGTAAAGGA 1610
QY 461 AlaProSerProLysValGluThrPyrArgGluGlyThrLeuLeuGluAspSerProAsp 480
Db 1611 GTCCCATCTCCCTAAGGTTGAGTGGTATAGAGAAGGACCTTTAATAGAAGATTCTCCAGAT 1670
QY 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluPro 494
|||||
Db 1671 TTTAGGATTTTACAGAAAAACCTCGATCCATGGCAGAGCCA 1712
RESULT 2
AAS74798
ID AAS74798 standard; cDNA; 1228 BP.
AC AAS74798;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #10602.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR
DR P-PSDB; ABG10611.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 10602; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1228 BP; 323 A; 319 C; 314 G; 272 T; 0 other;

Alignment Scores:
Pred. No.: 2,46e-100 Length: 1228
Score: 2164.00 Matches: 407
Percent Similarity: 99.76% Conservative: 1
Best Local Similarity: 99.51% Mismatches: 0
Query Match: 31.41% Indels: 1
DB: 23 Gaps: 0
US-09-818-990B-2 (1-1320) x AAS74798 (1-1228)
QY 913 MetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspGluIle 932
Db 1 ATCAATGAAATAGAGTTTCGCTTGAACGCTACTCTGTTGATGAATCAGATGAAATT 60
QY 933 GlnHisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLys 952
Db 61 CAACATGATGATCCCGGCAAGTGTATTGCTCCCATCTTTGACAAGAGACTCAAG 120
QY 953 HisPheArgValThrGluGlySerProValThrPheThrCysLysIleValGlyIlePro 972
Db 121 CACTTCGGGTACAGAAGCTCTCCAGTTACATTACCTGCAAAATTTGTTGGGATACCT 180
QY 973 ValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCys 992
Db 181 GTTCCAAAGGTTTACTGGTTCAAAGATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGC 240
QY 993 LysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAsp 1012
Db 241 AAAATGAGCGAGAGAGATGGGACATGCTCTCTGCACATTGAATCCACTACAGTGAT 300
QY 1013 AspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGly 1032
Db 301 GACGATGGCACTACACCATCATGGCAGCACACCCAGGGGAGAAATCACTGTTCTGGC 360
QY 1033 HisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis 1052
Db 361 CACTTGATGTACAAAGTTTGTCCCATTCGCAGTCGGCTAACCTCTGCTGCTCAGTCTCAC 420
QY 1053 ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArg 1072
Db 421 AGGGGAAGATCCCGAGTGCAGAAAGAGAGACAAAGAGCCCTACAGGAACGCTTTTCCGA 480
QY 1073 ProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeu 1092
Db 481 CCACATTTCTCGAGGCTCTCGGGATATGCTAGCTCATGAGGGGCCCTCTGTGGCTG 540
QY 1093 AspCysLysValSerGlyLeuProProGlu-LeuThrTrpLeuLeuAsnGlyGlnPr 1112
Db 541 GACTGTAAGGTGAGTGGTTTACCGCCCGGAGCATGACATGGCTACTCAATGGCCAACC 600
QY 1112 oValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLe 1132
Db 601 TGTGCTACCAGATGCTCCACAAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCT 660
QY 1132 uIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysTh 1152
Db 661 CATTGACCCCTCACTCAGCGCGCAGCGGACCTATAAGTGTGATCGCTACCAACAAAC 720
QY 1152 rGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaPr 1172
Db 721 CGGCGAGAATTCITTTAGTCTGGAGCTCTCTGTAGTAGCCAAAGAGGTGAAGAAAGCACC 780
QY 1172 oValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGl 1192
Db 781 TGTGATCTCTGGAGAAACTACAGAACTCGGTGTTCCGGAAGGCCACCCCGTGGAGACTGGA 840
QY 1192 uCysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGluThrIl 1212
Db 841 GTCCGGCGGTATAGGCATGCCACCTGTGTTCTTACTGGAAGAAAGACAAATGAGACAT 900
QY 1212 eProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLe 1232
Db 901 CCCTTGCAACGAGAGAGGATCAGTATGCACCAGGACACAAACAGGGGTATGCCTGCCCTTCT 960

```
Qy 1232 uileGlnProAlaLysSerAspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAl 1252
|||||
Db 961 CATTACGCCAGCCAAAGAAATCAGACGCTGGATGTCACACGTTGTCAGCCCAAGATGAAGC 1020

Qy 1252 aGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisGlnIlePr 1272
|||||
Db 1021 CGGCATCGTGTGTCGACACTGCCAGGCTGGATATATACGCTCAGTGGCACCATCAGATCCC 1080

Qy 1272 oProProMetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysGlyLe 1292
|||||
Db 1081 ACCGCCCATGTCCTCGCGCCAGTGGCAGTCGCTACGGATCTCTCACCAAGTAAGGACT 1140

Qy 1292 uAspIlePheSerAlaPheSerMetGluSerThrMetValTyrSerCysSerSerAr 1312
|||||
Db 1141 TGACATATTTCTGCTCTTCTCCATGGAAGCAGCATGGTGTATTTCATGCTCTCTCG 1200

Qy 1312 gSerValValGluSerAspGluLeu 1320
|||||
Db 1201 GAGTGTAGTGAGAGTGAACATT 1225

RESULT 3
AAC98221
ID AAC98221 standard; cDNA; 1823 BP.
AC AAC98221;
XX
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence S8Q ID NO:231.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN W020005351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR P-PSDB; AAB53464.
XX
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
XX
PS Claim 1; Page 653; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnery, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
```

CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX
SQ Sequence 1823 BP; 540 A; 452 C; 437 G; 387 T; 7 other;

Alignment Scores:

Pred. No.: 2,56e-67 Length: 1823
Score: 1510.50 Matches: 293
Percent Similarity: 69.09% Conservative: 87
Best Local Similarity: 53.27% Mismatches: 118
Query Match: 21.92% Indels: 52
DB: 21 Gaps: 7

US-09-818-990B-2 (1-1320) x AAC98221 (1-1823)

```
Qy 818 SerProThrSerArgIle-----GlnAsnProValAlaPheLeuSerSerValLeu 834
||||| :||| ||| ||||| ||||| :|||
Db 2 TCGCCTGCCACCGCTTCGGCCAGCGCCAGCGCCGCTTCCTCAGCGCTCTGCTG 61

Qy 835 ProSerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSer 854
||||| ||| ||||| ||||| |||
Db 62 CCCTCGCAGCG-----CGCG-----GNCGGGTCAACGCCCTGCCCAAG 109

Qy 855 MetProSerGlnGlyLeuAla-LysLysAsnThrLysSerProGlnProValAsnAspAs 874
::: :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 110 GTGTACCCCGCGAGATTCCAAAGAAGCGCAGTAGAACTGTAGATACCTCCCGATGA 169

Qy 874 pAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSe 894
: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||
Db 170 GGAATTCAAGGACAAAGGATGCTGTATTCAAGACCTGGAACGAAACTTCGCTTCAA 229

Qy 894 rAspValArgProAsn----- 899
::: |||
Db 230 GGAGGACCTCTCTGAACAATGGCCGCGAGGTTAAACATACGAAGAAGATGGCTCGTCG 289

Qy 899 ----- 899

Db 290 ACTGCTAGGTGCTGACAGTCCAACTGCTTTAATATTACGAGCCAGAGGAAACAGC 349

Qy 900 -GlnGlnGluTyrLysLysSerSerPheGluGlnArgLeuMetAsnGluIleGluPheAr 919
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 350 TAATCAGGAATACAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATAGAGTACAG 409

Qy 919 gLeuGluArgThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProTh 939
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 410 GCTAGAAAGTCTCTGTGGATGAATCAGGTGATGAAGTTTCAGTATGGAGATGGCTGT 469

Qy 939 rGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGl 959
::: ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 470 GGAATGGAATGGCACCATTCTTTGAGATGAAGCTGAAACATTACAGATCTTTGAGGG 529

Qy 959 ySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPh 979
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 530 AATGCCAGTAACCTTCACATGTAGAGTGGCTGGAATCCAAAGCCAAAGATCTATTGGTT 589

Qy 979 eLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAs 999
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 590 TAAAGATGGGAAGCAGATCTCTCCAAGAGTGTATCATACACCATCAAGAGATCTCGA 649

Qy 999 pGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrI 1019
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 650 TGGGACCTGCTCCCTCCATACCACAGCCTCCACCTAGATGATGATGGGAATTATACAAT 709

Qy 1019 eMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLe 1039
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 710 TATGGCTGCAAAACCTCAGGCGCGCATCAGTTGTACTGGACGGCTAATGTTACAGGCTGT 769

Qy 1039 uProIleArgSerArgLeuThrSerAlaGlyGlnSerHis-----Ar 1053
: ||| ||| :|||
Db 770 CAACAAAGAGGTGGAAGTCCCGGTCCTCCCTCAGGCGCATCTCATGTCAGAAAGGCGCTCG 829
```


Qy	921	GluArgThrProValAspGluSerAspAspGluLeuGlnHisAspGluLeuProThrGly	940
Db	583	GAAGAGTCTCCTGGATGAATCAGGTATGAGTAATCAAGTATGAGAGATGTCCTGTGGAA	642
Qy	941	LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer	960
Db	643	AACGGAATGGACCATCTTTGAGATGAAGCTGAACCATTAACAGATCTTTGAGGGGAATG	702
Qy	961	ProValThrPheThrCysLysIleValGluIleProValProLysValTyrTrpPheLys	980
Db	703	CCAGTAACTTTTCATGTAGAGTGGCTGGAAATCCAAAGCCAAAGATCTATTGGTTTAAA	762
Qy	981	AspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGly	1000
Db	763	GATGGGAAGCAGATCTCTCCAAAGAGTGATCACTACACCATTTCAAAGAGATCTCGATGGG	822
Qy	1001	ThrCysSerLeuHisIleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMet	1020
Db	823	ACCTGTCTCCTCCATACCAAGCCTCCACCTTAGATGATGATGGGAATATATACAATTATG	882
Qy	1021	AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro	1040
Db	883	GCTCCAAACCTCAGGCCGCATCACTGTCTGACGGCTAAATGGTACAGGCTGTCTCAAC	942
Qy	1041	IleArgSerArgLeuThrSerAlaGlyGlnSerHis-----ArgGly	1054
Db	943	CAAAAGGTGCGAAGTCCCGGTCTCCCTCAGGCCATCTCATGTGCAAGAGCCCTCGTTCT	1002
Qy	1055	ArgSerArgValGlnGluArgaspLysGluProLeuGlnGluArgPheArgProHis	1074
Db	1003	AGATCAAGGACAGTGGAGACGAAATGAACCAATTCAGGAGCGATTTCTCAGACCTCAC	1062
Qy	1075	PheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCys	1094
Db	1063	TTCTTGCAGGCTCTGGAGATCTGACTGTTCAAGAAGAAACAACTCGCAAGATGGACTGC	1122
Qy	1095	LysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeu	1114
Db	1123	AAAGTCAGTGGTTTACCAACCCAGATCTAAGCTGCAACTAGATGGAAAGCCCGTACGC	1182
Qy	1115	ProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAsp	1134
Db	1183	CCTCACAGTCTCACAGAATGCTGGTGGCTGAGAACGGGGTGCATCTCTGATCATAGAG	1242
Qy	1135	ProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGln	1154
Db	1243	CCAGTCAGTCAGTGATGCCGCATCTACACATGTATAGTACCACCGAGAGGACAG	1302
Qy	1155	AsnSerPheSerLeuGluLeuSerValAlaLysGluValLysLysAlaProValIle	1174
Db	1303	AACTCATTCAGCTGGAGCTTGGTGTGCTGCTAAAGAAGCACACAAACCCCTGTGTTT	1362
Qy	1175	LeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArg	1194
Db	1363	ATTGAGAAGCTCCAAACACAGAGTGTGCTGATGGGTACCCAGTGGCGCTGGAATGTCGT	1422
Qy	1195	ValIleGlyMetProProProValPheTyrTrpLysLysAspAsnGluThrIleProCys	1214
Db	1423	GTATTGGAGTGGCACCACTCAGATATTTTGGAGAAGAAATGAATCACTCACTCAC	1482
Qy	1215	ThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGln	1234
Db	1483	AGCACTCACCGAGTGAGCATGCACCAGGACACCACGGCTACATCTGCCCTCATTCAG	1542
Qy	1235	ProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIle	1254
Db	1543	GGAGCCACAAAAGAAGATGCTGGTGGTATCTGTGTGTGAGCAAGAATGAAGACGAGGATT	1602
Qy	1255	ValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis-----GlnIlePro	1272
Db	1603	GTGTCTGTACTGCCAGGTGGACGTTTTACACCCAGTGGCATCACTACAGCAGCAC	1662
Qy	1273	ProProMetSerValArqProSerGlySerArqTyrGlySerLeuThrSerLysGlyLeu	1292

Db	1663	AAGCCAAAAAAGTACGGCCTCAGCCAGTCGCTATGCAGCACTTTCGGACCGAGGACTA	1722
Qy	1293	AspIlePheSerAlaPheSerSerMetGluSerThrMetValTySerCysSerSerArg	1312
Db	1723	GACATCAACGAGCGTTCCAACT--GAGGCCAACCCATCTCACCTGACACTGATACT	1779
Qy	1313	SerValValGluSerAspGluLeu	1320
Db	1780	GCCTTGGTAGAAAGTGAGGACCTG	1803
RESULT 5			
ABA06552			
ID	ABA06552	standard; cDNA; 2059 BP.	
AC	ABA06552;		
XX			
DT	10-JAN-2002	(first entry)	
DE	Human cDNA SEQ ID NO: 218.		
KW	Human; gene therapy; neural disorder; immune system disorder;		
KW	muscular disorder; reproductive disorder; gastrointestinal disorder;		
KW	pulmonary disorder; cardiovascular disorder; renal disorder;		
KW	proliferative disorder; inflammation; ss.		
OS	Homo sapiens.		
XX			
PN	WO200154474-A2.		
XX			
PF	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01349.		
XX			
PR	31-JAN-2000; 2000US-179065P.		
PR	04-FEB-2000; 2000US-180628P.		
PR	24-FEB-2000; 2000US-184664P.		
PR	02-MAR-2000; 2000US-186350P.		
PR	16-MAR-2000; 2000US-189874P.		
PR	17-MAR-2000; 2000US-190076P.		
PR	18-APR-2000; 2000US-198123P.		
PR	19-MAY-2000; 2000US-205515P.		
PR	07-JUN-2000; 2000US-209467P.		
PR	28-JUN-2000; 2000US-214886P.		
PR	30-JUN-2000; 2000US-215135P.		
PR	07-JUL-2000; 2000US-216647P.		
PR	07-JUL-2000; 2000US-216880P.		
PR	11-JUL-2000; 2000US-217487P.		
PR	11-JUL-2000; 2000US-217496P.		
PR	14-JUL-2000; 2000US-218290P.		
PR	26-JUL-2000; 2000US-220963P.		
PR	26-JUL-2000; 2000US-220964P.		
PR	14-AUG-2000; 2000US-224518P.		
PR	14-AUG-2000; 2000US-224519P.		
PR	14-AUG-2000; 2000US-225213P.		
PR	14-AUG-2000; 2000US-225214P.		
PR	14-AUG-2000; 2000US-225267P.		
PR	14-AUG-2000; 2000US-225268P.		
PR	14-AUG-2000; 2000US-225270P.		
PR	14-AUG-2000; 2000US-225447P.		
PR	14-AUG-2000; 2000US-225757P.		
PR	14-AUG-2000; 2000US-225758P.		
PR	14-AUG-2000; 2000US-225759P.		
PR	18-AUG-2000; 2000US-226279P.		
PR	22-AUG-2000; 2000US-226681P.		
PR	22-AUG-2000; 2000US-226686P.		
PR	22-AUG-2000; 2000US-227182P.		
PR	23-AUG-2000; 2000US-227009P.		
PR	30-AUG-2000; 2000US-228924P.		
PR	01-SEP-2000; 2000US-229287P.		
PR	01-SEP-2000; 2000US-229343P.		
PR	01-SEP-2000; 2000US-229344P.		

```
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-233066P.
PR 21-SEP-2000; 2000US-234233P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-246617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 05-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-476161/51.
XX P-PSDB; ABB10330.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX Claim 1; SEQ ID NO: 218; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
SQ Sequence 2059 BP; 589 A; 460 C; 502 G; 508 T; 0 other;

Alignment Scores:
Pred. No.: 2.08e-61 Length: 2059
Score: 1394.00 Matches: 261
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.62% Mismatches: 0
Query Match: 20.23% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-2 (1-1320) x ABA06552 (1-2059)
QY 1059 GlnGluArgaspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAla 1078
Db 7 AGAGAAAGAGACAAAGAGCCCTACAGGACGCTTTTCCGACACACATTTCTCGAGGCT 66
QY 1079 ProGlyAspMetValAlaHisGluGlyArgLeuGlyArgLeuGlyArgLeuGlyValSerGly 1098
Db 67 CCTGGGATATGGTAGCTCATGAGGGGGCCCTCTGTGGCTGGAGTGAAGTGAGTGGT 126
QY 1099 LeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSer 1118
Db 127 TTACCGCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTCTACACAGATGCTCC 186
QY 1119 HisLysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuLeuLeuLeuThrGln 1138
Db 187 CACAAGATGTGTGTGTCAGGAGCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAG 246
QY 1139 ArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSer 1158
```

```
Db 247 CGCGACGCGAGCTATAGTCATCGCTACCAACAAACCGGCGAGAAATCTTTAGT 306
Qy 1159 LeuGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeu 1178
Db 307 CTGGAGCTCTCTAGTAGCAAGAGGTGAAGAACACCTGTGATCCTCGGAAACTA 366
Qy 1179 GlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMet 1198
Db 367 CAGAACTGCGGTGTTCGGAAGGCCACCCCGTGAGACTGGAGTCCGCGTGATAGGCATG 426
Qy 1199 ProProProValPheTyrTrpPLeLysAspAsnGluThrIleProCysThrArgGluArg 1218
Db 427 CCCCCACCTGTGTCTACTGGAAGAAACAAATGAGACCATCCCTTGCCACGAGAGAGG 486
Qy 1219 IleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLys 1238
Db 487 ATCAGTATGCAACAGGACACAAACAGGTATGCCCTTCTCATTCAGCCAGCCCAAGAAA 546
Qy 1239 SerAspAlaGlyTyrTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThr 1258
Db 547 TCACAGCGTGGATGGTACACGCTTCTCAGCCAAGAAATGAGCCGCATCGTCTGTGCACT 606
Qy 1259 AlaArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProprometSerValArg 1278
Db 607 GCCAGGCTGGATATATACGCTCAGTGGCACCACCATCAGATCCACCGCCCATCTCTGTCCGG 666
Qy 1279 ProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPhe 1298
Db 667 CCCAGTGGCAGTCGCTACGGATCTCTCACCAGTAAGGACTTGACATATTTTCGCCCTTT 726
Qy 1299 SerSerMetGluSerThrMetValTyrSerCysSerSerArgSerValGluSerAsp 1318
Db 727 TCCTCCATGGAAACGACGATGCTGTATTCATGCTCTCTCGGAGTGTAGTGGAGATGAT 786
Qy 1319 GluLeu 1320
Db 787 GAACCT 792

RESULT 6
AAS28777
ID AAS28777 standard; cDNA; 2232 BP.
XX
AC AAS28777;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immunoglobulin encoding cDNA SEQ ID No 23.
XX
KW Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
PN W0200155315-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01326.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
```

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0241617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Barash SC, Ruben SM;

WPI: 2001-457725/49.

P-PSDB; AAU17989.

Isolated novel immunoglobulin polypeptide for monitoring the presence and progression of diseases and for diagnosis -

Claim 1: SEQ ID No 23; 551pp; English.

XX Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the immunoglobulin polypeptides of the invention, and primers for the polynucleotides. The polynucleotides and polypeptides can be used to diagnose a pathological condition or a susceptibility to a pathological condition in a subject by determining the presence or absence of a mutation in a DNA sequence or determining the presence or amount of expression of the protein. Alternatively the identification of a binding

CC partner to a sequence allows determination of changes in protein activity. The sequences can be used as research tools for receptors or other signal transduction pathway proteins that interact with the polypeptides of the invention and can be used to treat, prevent or diagnose various types of disorders such as neurological disorders, cardiovascular disorders, gastrointestinal disorders, reproductive disorders, immune system disorders, renal disorders, muscular disorders, pulmonary disorders, proliferative disorders and cancer.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 2232 BP; 617 A; 508 C; 528 G; 569 T; 10 other;

Alignment Scores:

Pred. No.:	1.15e-60	Length:	2232
Score:	1380.00	Matches:	239
Percent Similarity:	99.23%	Conservative:	0
Best Local Similarity:	99.23%	Mismatches:	2
Query Match:	20.03%	Indels:	0
DB:	22	Gaps:	0

US-09-818-990B-2 (1-1320) x AAS28777 (1-2232)

QY	1060	GluArgAspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaPro	1079
DB	3	GAAAGAGACAAAGAGCCCTACAGAAACGCTTTTCGGACCACATTTCTCGAGGCTCCT	62
QY	1080	GlyAspMetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeu	1099
DB	63	GGGGATATGGTAGTCATCAGGGGGCGCTCTGTCGGCTGGACTGTAAGGTGAGTGGTTA	122
QY	1100	ProProProGluLeuThrTrpLeuLeuAsnGlyGluProValLeuProAlaSerHis	1119
DB	123	CCGCCCCCGAGCTGACATGGCTACTCAATGGCCACCTGTGTACAGATGCTCCAC	182
QY	1120	LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArg	1139
DB	183	AGATGCTGTGTAGGAGACCCGGAGTCCACTCTCTGCTATTGACCCACTCCTACGGCG	242
QY	1140	AspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu	1159
DB	243	GACGAGGGACCTATAAGTGCATCGCTACCAACAAACCGGCGAGAATCTTTTAGTCTG	302
QY	1160	GluLeuSerValValAlaLysGluValLysLysAlaProValLleLeuGluLysGln	1179
DB	303	GAGCTCTGTAGTACCAAGAGGTGAAGAAAGCACTGTGATCTCTGGAGAATACAG	362
QY	1180	AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro	1199
DB	363	AACTGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCC	422
QY	1200	ProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle	1219
DB	423	CCACCTGTGTCTACTGGAGAAAGACAATGAGACCATCCCTTTGCACAGAGAGGATC	482
QY	1220	SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSer	1239
DB	483	AGTATGACACAGGACACACAGGGTATGCTGCTCTCTCATTCAGCCAGCAAGAAWCA	542
QY	1240	AspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla	1259
DB	543	GACGCTGGATGGTACAGTTGTGAGCAAGATGAAGCGGCATCGTGTCTGCACATGCC	602
QY	1260	ArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProMetSerValArgPro	1279
DB	603	AGCTGGATATACGCTCAGTGGCCACCATCAGATCCACCGSCCATGTCTGTCGGGCC	662
QY	1280	SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSer	1299
DB	663	AGTGGCAGTCGCTACCGATCTCTCACCAGTAAGAGGATTGACATATTTCTGCTTTTC	722
QY	1300	SerMetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSerAspGlu	1319

|||||
Db 723 TCCATGAAACGACGATGCTGTTTCATGCTCTTCTCGGAGTGTAGTGAGATGATGAA 782
Qy 1320 Leu 1320
Db 783 CTT 785
RESULT 7
ABA06731
ID ABA06731 standard; cDNA; 702 BP.
XX
AC ABA06731;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 397.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225759P.
PR 14-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226686P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 05-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.

PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-476161/51.
DR P-PSDB; ABB10509.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX Claim 1; SEQ ID NO: 397; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
XX Sequence 702 BP; 179 A; 195 C; 187 G; 137 T; 4 other;
SQ

Alignment Scores:
Pred. No.: 7.85e-49 Length: 702
Score: 1133.50 Matches: 226
Percent Similarity: 96.17% Conservative: 0
Best Local Similarity: 96.17% Mismatches: 9
Query Match: 16.45% Indels: 4
DB: 22 Gaps: 0

US-09-818-990B-2 (1-1320) x ABA06731 (1-702)

QY 1060 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 1079
|||||
Db 3 GAAAGAGACAAAGAGCCCTACAGGACGCTTTTCCGACACATTTCTCGAGGCTCT 62
|||||
QY 1080 GlyAspMetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeu 1099
|||||
Db 63 GGGGATATGGTAGCTCATGAGGGGCGCTCTGTGGCTGGAGTGAAGGTGAGTGGTTA 122
|||||
QY 1100 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 1119
|||||
Db 123 CG-CCCCGGAGCTGACATGGCTACTCAATGGCCACCTGTGTACCATGCTCTCCAC 181
|||||
QY 1120 LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArg 1139
|||||
Db 182 AAGATGCTGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCGC 241
|||||
QY 1140 AspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 1159
|||||
Db 242 GACGA-GGGACCATATAAGTCATCGCTTACCAACAAACCGGGCAGAAATCTTTAGTCTG 300
|||||
QY 1160 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGln 1179
|||||

Db 301 GAGCTCTCTAGTAGCCAAAGAGGTGAAGAAAGACACCTGTGATCCTGGAGAACTACAG 360
|||||
QY 1180 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 1199
|||||
Db 361 AACTGGCGTGTCCCGAAGGCCACCCCGTGAGACTGGAGTCCGCGGTGATAGCATGCC 420
|||||
QY 1200 ProProValPheTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle 1219
|||||
Db 421 CCACCTGTGTCTACTGGAAGAAGACATGACACCATCCCTTGACACAGAGAGGATC 480
|||||
QY 1220 SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuLeuGlnProAlaLysLysSer 1239
|||||
Db 481 AGTATGCACCCAGGACACACAGGATGCTGCTCTTCATTGACCGCAGCAAGAAATCA 540
|||||
QY 1240 AspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla 1259
|||||
Db 541 GACGCTGGATGGTACACGTTGTGAGCCAAAGATGAAGCCGCGCATCGNGTCTGCTGCACTGCC 600
|||||
QY 1260 ArgLeuAspIleTyrAlaGlnTrpHisHisGlnIleProProProMetSerValArgPro 1279
|||||
Db 601 AGGCTGGATATATACGCTCAGTGGCACCACATCATCCACSGCCATGTC--TGCCGGCCCC 658
|||||
QY 1280 SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIle 1294
|||||
Db 659 AGNGGCAGTCGNTACGGATCTCTAACAGT-AAAGGACTTGACATA 702
|||||
RESULT 8
AAS28841
ID AAS28841 standard; cDNA; 702 BP.
XX
XX AAS28841;
XX
XX 07-NOV-2001 (first entry)
XX Human immunoglobulin encoding cDNA SEQ ID No 87.
DE
XX Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
XX Homo sapiens.
XX
XX WO200155315-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01326.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231142.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231144.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231144.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233066.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457725/49.

P-PSDB; AAU18053.

Isolated novel immunoglobulin polypeptide for monitoring the presence and progression of diseases and for diagnosis -

Claim 1; SEQ ID No 87; 551pp; English.

Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the immunoglobulin polypeptides of the invention, and primers for the polynucleotides. The polynucleotides and polypeptides can be used to diagnose a pathological condition or a susceptibility to a pathological condition in a subject by determining the presence or absence of a mutation in a DNA sequence or determining the presence or amount of expression of the protein. Alternatively the identification of a binding partner to a sequence allows determination of changes in protein activity. The sequences can be used as research tools for receptors or other signal transduction pathway proteins that interact with the polypeptides of the invention and can be used to treat, prevent or diagnose various types of disorders such as neurological disorders, cardiovascular disorders, gastrointestinal disorders, reproductive disorders, immune system disorders, renal disorders, muscular disorders, pulmonary disorders, proliferative disorders and cancer.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 702 BP; 179 A; 195 C; 187 G; 137 T; 4 other;

Alignment Scores:

Pred. No.: 7, 85e-49 Length: 702
Score: 1133.50 Matches: 226
Percent Similarity: 96.17% Conservativity: 0
Best Local Similarity: 96.17% Mismatches: 9
Query Match: 16.45% Indels: 4
DB: 22 Gaps: 0

US-09-818-990B-2 (1-1320) x AAS28841 (1-702)

QY 1060 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 1079
DB 3 GAAAGACAAAGAGCCCTACAGGACGCTTTTCCGACACATTTCTCGAGGCTCT 62
QY 1080 GlyAspMetValAlaHisGluGlyArgLeuGlnGlnGlnGlnGlnGlnGlnGln 1099
DB 63 GGGGATATGGTAGCTCATGAGGGGCGCTCTGTGGGCTGGACTGTAAGGTGAGTGGTTTA 122
QY 1100 ProProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 1119
DB 123 CG-CCCCGGAGTGCATGGCTACTCAATGGCCAACTGTGTACAGATGCTCCAC 181
QY 1120 LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuLeuLeuLeuLeu 1139
DB 182 AAGATGCTGTGAGGAGACGGGAGTCCACTCTGTCTCATTTGACCCACTCACTACGCGC 241
QY 1140 AspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 1159
DB 242 GACGA-GGGACCTATAAGTGCATGCTTACCAACAAACGGGCGAGAATCTTTTAGTCTG 300
QY 1160 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGln 1179
DB 301 GAGCTCTGTAGTAGCCAAAGAGGTGAAGACCACTGTGTATCTTGGAGAACTACAG 360
QY 1180 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 1199
DB 361 AACTGCGGTGTCCGGAAGGCCACCCCGTGAGACTGGAGTGCCGCGTATAGGATGCC 420
QY 1200 ProProValPheTrpTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle 1219
DB 421 CCACCTGTGTCTACTGGGAAGAAGCAATGAGACCATTCCCTTGCACACAGAGAGATC 480
QY 1220 SerMetHisGlnAspThrTrpGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSer 1239
DB 481 AGTATGACCCAGGACACACAGGGTATGCTGCTCTCTCATTCAGCCAGCCCAAGATCA 540
QY 1240 AspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla 1259
DB 541 GACGCTGGATGGTACAGCTTGTACGCCAAGAAATGAAGCCGATCGNGTGTGCACTGCC 600
QY 1260 ArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProProMetSerValArgPro 1279
DB 601 AGGCTGGATATATACGCTCAGTGGCCATCAGATCCAGCGCCATGTC--TGCCGGGCC 658
QY 1280 SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIle 1294
DB 659 AGNGGCGAGTCGNTACGATCTCTAACAGT-AAAGACTTGACATA 702

RESULT 9
AAH04212

ID AAH04212 standard; cDNA; 855 BP.

XX AC AAH04212;

XX DT 26-JUN-2001 (first entry)

XX XX Human cDNA clone (5'-primer) SEQ ID NO:1047.

XX DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX KW Homo sapiens.

XX OS

XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 1: SEQ ID 1047; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences. AA92446 to
XX AA93893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX SQ Sequence 855 BP; 257 A; 214 C; 176 G; 205 T; 3 other;
Alignment Scores:
Pred. No.: 3, 88e-38 Length: 855
Score: 922.50 Matches: 185
Percent Similarity: 95.45% Conservativity: 4
Best Local Similarity: 93.43% Mismatches: 9
Query Match: 13.39% Indels: 1
DB: 22 Gaps: 0
US-09-818-990B-2 (1-1320) x AAH04212 (1-855)
QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
DB 231 ATGCAACACGACACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 290
QY 21 LeuAlaGluThrArgHisArgGlyAsnGluArgSerArgAlaGluProSerSerAsn 40
DB 291 TTACGTGAACACACACATCGGGGAAACAATGAGAGGAGTGCAGCGGAGCCCTCTCCCAAC 350
QY 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlyGlyGlyGly 60
DB 351 CTTGCCATTTTCGGCAGTCTCTGGGCGCGCTGAAGGAGCGGAGGCGCAAGATGACCTT 410

Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluGluLeuAspGluSerValAsnLeuAlaArg 80
Dbb 411 CCAGATCTTTAGCCCTTTCTGAGCCAAGAAGAAATAGACGAAAGTGTCAATTGGCAAGA 470
Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Dbb 471 CTGGCCATCAATTAGACCCCTTTGGAGAGGAGGAGATGAAGCTCAAGCTAGAAAAGACTT 530
Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Dbb 531 TCTCTGATCAGATGAACATCCTCAATTAAGTTTGGACCTPACTTCTGCCAGAT 590
Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Dbb 591 AACCTCGAAGTCCACCACTCTTAAGAAAGCCCGAGGAGGCAAAAGGCCACAGTAT 650
Qy 141 CysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Dbb 651 TGGTCTGAACCCAGTCCAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGAAGAG 710
Qy 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
Dbb 711 CTATCTCTCT--TTCAATCCACAGCTTCCAAAGGATTAGACCTCGTGCCTGCNAARAC 768
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSer 198
Dbb 769 CNCAGAGTNAACTGGAATCTCAAAACAAAGTATGCAGGAAACAGCTCAGTTCT 822
RESULT 10
ABK64829
ID ABK64829 standard; DNA; 81940 BP.
XX
AC ABK64829;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human benign prostatic hyperplasia gene #724.
XX
KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
OS Homo sapiens.
XX
PN WO200212440-A2.
XX
PD 14-FEB-2002.
XX
PF 07-AUG-2001; 2001WO-US24708.
XX
PR 07-AUG-2000; 2000US-22323P.
PR 05-JUN-2001; 2001US-0873319.
XX
PA (GENE-) GENE LOGIC INC.
PA (NTSB) JAPAN TOBACCO INC.
XX
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
XX
XX WPI; 2002-257476/30.
XX
XX
PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by
PT detecting expression levels of one or more genes in prostate cells from
PT patient that are differentially regulated compared to normal prostate
PT cells -
XX
XX Disclosure; Page 405-429; 444pp; English.
XX
XX The invention relates to a method of diagnosing (I) the onset or
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for
CC or identifying an agent that modulates the onset or progression of BPH.
CC The method is based on changes in gene expression in BPH tissue isolated
CC from patients exhibiting different clinical states of prostate
CC hyperplasia as compared to normal prostate tissue. (I) comprises
CC detecting the expression levels of one or more genes in prostate cells

CC from the subject that are differentially regulated compared to normal
CC prostate cells. (II) comprises preparing a first gene expression profile
CC of BPH cells or BPH-like cell population, exposing the cells to the
CC agent, preparing a second gene expression profile of the agent exposed
CC cells, and comparing the first and second gene expression profiles.
CC (I) is useful for diagnosing the onset or progression of BPH. (II) is
CC useful for identifying an agent that modulates the onset or progression
CC of BPH. The methods are useful to present information identifying
CC the expression level in a tissue or cells, by comparing the expression
CC level of genes given in the specification in the tissue or cells to the
CC level of expression of gene in the database, and displaying the
CC expression levels of at least one gene in the tissue or cell sample
CC compared to the expression level in BPH. Agents using (II) are useful for
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human
CC benign prostatic hyperplasia gene sequences of the invention.

XX Sequence 81940 BP; 26373 A; 17097 C; 18901 G; 19569 T; 0 other;

Alignment Scores:

Pred. NO.:	2.38e-30	Length:	81940
Score:	810.50	Matches:	360
Percent Similarity:	34.15%	Conservative:	228
Best Local Similarity:	20.91%	Mismatches:	512
Query Match:	11.76%	Indels:	622
DB:	24	Gaps:	54

US-09-818-990B-2 (1-1320) x ABK64829 (1-81940)

Qy	100	LeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGln	119
Dbb	82	CTTAGTCCCAACCTTTTATAGGCACCTTTGAGAAATCAGAGTG-----	123
Qy	120	AspAsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGln	139
Dbb	124	-----CCTAGAAAGATGACAACT-----CAAGCACCAGC	153
Qy	140	TyrCysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGlu	159
Dbb	154	TTTACGACGCGTTACAAAGCGTTGTGTA-----CTGGAG	189
Qy	160	GluLeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLys	179
Dbb	190	GGTAGTACCGCAACCTTTGAGGCTCACATATTAGTGGTTTCCAGTTCCTGAGGTGAGCTGG	249
Qy	180	AsnHisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPhe	199
Dbb	250	TTTAGGGATGCCAGGTGATTTC-----ACTTCCACTCTGCCCGGCGTCCACATCTCCTTT	306
Qy	200	SerAspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThr	219
Dbb	307	AGCGAT-----GGCCGCGCTAAACTGACGATCCCGCGGTGACT	345
Qy	220	ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGlu	239
Dbb	346	AAAGCCAACAGTGGCAGATATTCCTG-----	372
Qy	240	GlnAlaAlaSerGluAlaAlaGlyLysThrProGlySerSerProSerSerLeu	259
Dbb	373	-----AAAGCCCAATGGATCTGGACAAGGACTAGTACTGTGAGCTTCTC	420
Qy	260	TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGlu	279
Dbb	421	GTGAAGCTGAGACA-----GCACCACCACTTCGTTCAACGACTGCAGCAGATGACC	474
Qy	280	ValProGluGlyThrArgValGlnLeuAspCysIleValGlyIleProProProGln	299
Dbb	475	GTGAGACAAGGAAGCCAAAGTGAGACTCCAACTGAGAGTACTGGAATCCCTTAACCCCTGTG	534
Qy	300	ValArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGln	319
Dbb	535	GTGAAGTCTTACCGGGATGGAGCCGAATCCAGAGCTCCCTTGATTTCCAAATTTACAA	594
Qy	320	AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr	339

```
Db 595 GAAGGCGACCTCTACAGCTTACTGATTGCAGAGCATACCTGAGGACTCAGGACCTAT 654
    |||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
Qy 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
    ||| |||:|||||: ||| |||:|||||:|||||: |||:|||||: |||
Db 655 TCAGTAAATGCCCAATAGCGTTGGAGAGCTACTTCGACTGCTGAATTAAGTTGTTCAA 714
    ||| |||:|||||: ||| |||:|||||:|||||: |||:|||||: |||
Qy 360 Gly-----ValSerSerAspSerGluGlyAspProAsn 371
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 715 GGTGAAGAAGTACCTGCTAAAGACAAAGACAAATTTGTCGACTGCTCAGATCTCA 774
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 372 LysGluGluMetAsnArgIleGlnLys----- 380
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 775 GAATCAAGACAAACCCGAATGAAAGAGAGATTGAAGCCACCTTTGATGCCAGATCAATT 834
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 381 -----Pro 381
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 835 GCAACAGTTGAGATGGTCATAGATGGTGGCGCTGGGCAACAGCTGCCACATAAAACACCT 894
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 382 AsnGluValSerProThrThrSerAlaValIleProProAlaValProGlnAla 401
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 895 CCCAGGATTCTCCGAAGCCAAAGTCAAGATCCCAACACACCGCTATTTGCTGCCAAA 954
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 402 GlnHisLeuValAlaGlnProArgValAlaThrIleGlnCysGlnSerProThrAsn 421
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 955 GCACAGCTGGCTCGGCAGCAGTCCCATCGCCATAAGACACTCCCTTCCCGGTCAGA 1014
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 422 TyrLeuGlnGlyLeuAspGlyLysProIle----- 431
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1015 CAGTGGGGCAGCCAGCCCATCTCCGGTCAGGTCCGTCTCOAGCAGCAAGAATCTCC 1074
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 432 -----IleAlaAlaProValPheThrLysMetLeuGlnAsnLeuSer 445
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1075 ACATCCCCCATCAGGTCTGTGTAGTCTCCATTTGCTGCTAGAGACTCAGGCATCCACC 1134
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 446 AlaSerGluGlnLeuValPheGluCysArgValLysGlyAlaProSerProLys 465
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1135 GTGCCACAGTCTCTGAAGTG-----CCTCCCCCT--- 1164
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 466 ValGluTrpTyrArgGluGly-----Thr 473
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1165 -----TGGAAAGCAAGAGGGCTAGTGGCCCTCTCATCTGAGGCTGAGATGAGAGAGCA 1218
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 474 LeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGlu 493
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1219 ACGTGCACACCTTACTCAGATCAGCAGCAGACAGAGAGATGGGAGGAGATACGGTGC 1278
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 494 ProGluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1279 CAGGCAAGTGAACCATCAGTGGTGTGCGGGTGTGCGGCCAGTGTGCGGCCAGTGCT 1338
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 514 ThrCysThrAla-----SerAsnLys 520
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1339 AGCTACGACGACAGAGGTGTGGCCACTGGTGTCTAAAGAGGTGAACAGATGCTGACAAA 1398
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 521 TyrGlyThrValSerIle----- 527
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1399 AGTCCAGCTGTGGCACTGTGTGTCGCCGTGATGGCCAGGTGAGAGAACCAAGTG 1458
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 528 -----AlaGlnLeuHisValArgGly 534
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1459 ATCAGCGCTGTAGACAGACTGCTCAGAGGACCAACAGACTGCTGTGCATCCAACT 1518
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 535 AsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAla 554
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1519 GCTCAAGAACAGGTGAAGAAAGGAGCGAGAGAGACTGCTGTAACTAAGGTAGTAGTGCC 1578
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 555 Ala-IleGluProGlnProSerProProHisSerGluPro----- 567
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1579 GCCGATAGCCAGGAGCAAGAAATTAATCAAGAACCAAGAAATAATTACCACAAG 1638
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 567 ----- 567
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
```

```
Db 1639 CAAGAGCAGATGCACGTAACTCATGAGCAGATAAGAAAGAACTGAAAAAACATTTGTA 1698
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 568 -----ProSerValGluGlnProProLysProLys-----Le 578
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1699 CCAAGGTAGTAATTCGCGAGCTAAAG-----CCAAGAACAAAGAACTAGAAATTTCT 1752
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 578 uGluGlyValLeuValAsnHisAsnGlu----- 587
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1753 GAAGAAATTACTAGAACAAACAAAGTAAGTCAAGNAGCAATAANTAGAGAACTAGG 1812
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 588 -----ProArgSerSerSerArgIle-GlyL 596
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1813 AAAACAGTTGTACCTAAAGTCATAGTTGCCACACCCAAAGTCAAGAACAAAGATTTAGTA 1872
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 596 eu-----ArgValHisPhe-----AsnLeuProGluA 605
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1873 TCAAGAGGTAGAGAGGCAATTAATTAACAAAGAGAACAAAGTGCATAAATACTCAGGAGAAG 1932
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 605 spAspLysGlySer-Glu-----Ala 611
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1933 ATGAGAAGGAGGCGGAGAAACTGCCCTGTCTACAATAGCAGTTGCTACTGCTAAAGCC 1992
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 612 SerSerGluAlaGlyValValThrArgGlnThr----- 623
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 1993 AAGAACAAGAAACAATACTGAGAACTAGAGAAACTATGGCTACTAGACAAAGAACAAATC 2052
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 623 ----- 623
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2053 CAAGTTACCATGGAAGGTGGACGTTGGAAAAAGGCTGAAGCTGTAGCAACAGTTGTT 2112
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 624 -----ArgProAspSer***GlnGluArg 631
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2113 GCTGCAGTAGACCGCCGAGTCAGAGAGCCAGAGAGCCTGGGCATCTTTGAAGATCC 2172
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 632 PheAsnGlyGlnAlaThr-----LysThr 639
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2173 TATGCTCAGCAGACCACTTTGGAGTACGGATATAAGGAACGCATTTCCGCGCCAAAGTA 2232
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 640 ProGlu---ProSerPheProValLysGluPro-----Pro 650
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2233 GCTGAGCTCTCCCAACGTCAGCCTCAGAACCCACGCTGTCCCTAAAGCAGTCAAGCCT 2292
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 651 ProValLeuAlaLysProLys-----LeuAspSerThrGlnLeuGlnLeuHis 667
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2293 AGAGTAATCCAGGCTCTCTCTGAGACTATATCAAACTACTGATCAAAAGGAATGCAC 2352
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 668 AsnGlnValLeuLeuGluGln-----HisGlnLeu 677
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2353 ATATCATCACAGATCAAGAAAACTACAGATCTAACACGGAAAGATTAGTCCATGTGGAT 2412
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 678 GlnAsnProProProSerSerProLysGluPheProPhe***MetThrValLeuAsn--- 696
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2413 AAACGCCCCGACAGCTAGCCCTCAC-----TTTACTGTTTCAAAAATTTCTGTT 2463
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 697 -----SerAsnAlaProAlaValThrThrSer*** 707
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2464 CCTAAGACAGAACATGGATATGAGGCATCAATAGCCGGTAGTGTATTTGCCACATTACAA 2523
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 708 LysGlnValLysAlaProSerSerGln-----ThrPheSerLeuAlaArgProLys 724
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2524 AAAGAGTTGTGAGCCACATCTTCTGCTCAGAAGATCAACAAATCGGTGAAGGCTCTACT 2583
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 725 TyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSerSerSerProVal 744
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2584 GTG-----AAGCCCAAGTGAGACTAGAGTAAGGGCAGAGCCCAACCCCTTCCACAG 2634
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 745 PheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeu 764
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2635 TTCCCTTCGCTGACACACCCAGACTACTACAAG----- 2667
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Qy 765 ValSerHisProSerValGlnThrLysSerProGlyLeuSerIleGlnAsnGluPro 784
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
Db 2668 ---AGTGAAGCTGGCTTGAGGTGAAAAAGGAAGTAGGGGTGAGCATCACTGGCACACC 2724
    ||| |||:|||||: ||| |||:|||||: ||| |||:|||||: |||
```

Qy 785 Leu----- 785
Db 2725 GTCCGTGAAGAGCGCTTTGAAGTACTGCACGGACGCGAAGCAAGGTAAACAGAAACAGCA 2784
Qy 786 ---ProProGlyProThrGlu---ProThrProProProPheThrPheSer----- 800
Db 2785 AGAGTACCAGCAGCACTGTTGAAATTCCTGTATCTCCACCAACTTTTGGTCTCGGGCTTAAAA 2844
Qy 801 -----IleProSerGlyAsnGlnPheGlnProArgCys---ValSerProIlePro 816
Db 2845 AATGTGACTGTGCATAGAAAGGTGAATCTGTACCTTGGAGTCCACATCTCTGGATACCACCA 2904
Qy 817 ValSerProThr-----SerArgIleGlnAsn 825
Db 2905 ---TCCCGCAGACGTGACATGTGTACAGGGAAGAGTACCAAAATCGAAAGTTCCTCATGACITC 2961
Qy 826 ProValAlaPheLeuSerSerVal----- 833
Db 2962 CAGATAACCTTCCAGAGTGGAAATTGCTCGCTTATGATTCGCGAAGCATTTTCGCGAAGAC 3021
Qy 833 ----- 833
Db 3022 AGCGGGCGATTACTTGCAGTGTCTGTAATGAGGCTGGAACCGTCAGCACATCTCTGTAT 3081
Qy 833 ----- 833
Db 3082 CTGGCTGTGCAGGTGTCAGAAAGATTTGAAAGGAACACACAGCGGTGACTGAGAAATTT 3141
Qy 834 -----LeuProSerLeuProAlaIleProProThrAsnAlaMet 846
Db 3142 ACTCAGAAAGAGAAACGCTTTTGGAGTCAAGAGATGTGGTATGACTGTGACTGATACGCTC 3201
Qy 847 ***LeuProArgSerAlaProSerMetProSer----- 857
Db 3202 ACAGAGGAACAGCAGGCGCTGGAGAACCTGCCGGCGCTTACTTTATTACAAAACCAAGTG 3261
Qy 858 ---GlnGlyLeuAlaLeuLysAsnThr-----LysSerPro 868
Db 3262 GTCCAGAAACTGGTGGAGGTGGGAGCGTGGTGTGGATGCCAAGTGGCGGCAACCCA 3321
Qy 869 GlnPro----- 870
Db 3322 AAGCCCCATGTATACTGGAAAAAATCTGGTGTCTCTTAACCACTGGATACAGATACAAA 3381
Qy 870 ----- 870
Db 3382 GTGAGTTACAAACAAACCCGGTGAATGCAAGCTGGTGATTTCTATGACTTTTGTGCTGAT 3441
Qy 870 ----- 870
Db 3442 GATGCTGGAGAATACACTATTGTTGTCGCAATAGCATGGAGAAACTTCTGCATCTGCT 3501
Qy 870 ----- 870
Db 3502 TCCTTGTGAAGAGCTGATTATGAGTTACTGATGAAGTCCCAGCAGAAATGCTTTAT 3561
Qy 871 -----ValAsnAspAsnIleArgGluThr----- 879
Db 3562 CAGACAAAGTGACTGCATTTGTTCAAGAACCTGAAGTTGGAGAAACAGCACCTGGATT 3621
Qy 880 -----LysAsnAlaValIleArg---AspLeu 887
Db 3622 GTATACTCTGAGTATGAAAAAGAGTATGAAAAAGAACCAAGCCTTAATTTAGGAAGAAATG 3681
Qy 888 GlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGluTyrLysIleSerSer 907
Db 3682 GCCAAAGATACTGTAGTGGTCAGAACTTATGTAGAAGATCAGGAATTCATATTTCTTCC 3741
Qy 908 PheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrPro----- 924
Db 3742 TTTGNAGAGACACTATTAAAGAAATTTGAATATAGAATAATAAAGACTACATATTAGAGAA 3801

Qy 925 ---ValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCysIle 943
Db 3802 CTTCCTTGAAGAAGATGGAGAAAGAAAGATGCAGTGTGACATTTCTGAATCTGAAGCTGTT 3861
Qy 944 AlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProValThr 963
Db 3862 GAATCAGGATTTGATTTAAGAATCAAGAATATAGAATTTCTTGGGGGATGGGTGTCACT 3921
Qy 964 PheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGlyLys 983
Db 3922 TTTTCATTCGCAAGATCTCTGGATATCCATTACCAAGATGTCTTGTACAAAGATGGCAAG 3981
Qy 984 GlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSer 1003
Db 3982 CGCATC---AAACATGGAGAAAGATACCAAAATGGACTTTTTCACAAATGGCAGAGTAGT 4038
Qy 1004 LeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsn 1023
Db 4039 CTGGCTATACCTGTGTCTTCCAGAAAGATGAAGAAATCTACATGCAATTTTTCGCAAGCAAT 4098
Qy 1024 ProGlnGlyArgIleSerCysSerGlyHisLeuMetValGln----- 1037
Db 4099 ATTAAGGAATGCAATTTGCTCAGGGAAATTTGATGTGGAGCCTGCTGCACCACCTTGGGA 4158
Qy 1038 -----SerLeu----- 1039
Db 4159 GCTCCGACTTACATTTCCACACTAGAGCCAGTGCAGAGAATCAGATCTCTCTCCACGT 4218
Qy 1040 -----ProIleArg-----SerArgLeuThr 1046
Db 4219 TCAGTGCAGAGGTCTCCTATACGCATGTCTCTGCACGGATGCACCTGCAGAGATGTCT 4278
Qy 1047 SerAlaGlyGlnSerHisArgGlyArgSer-----ArgValGlnGluArgAspLys 1063
Db 4279 CCTGCAGGATGTCCTGCCAAGAAATGTCCTCGACTAGGCTGGAGGAGAGAGATGAG 4338
Qy 1064 GluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetVal 1083
Db 4339 TCACAACCT---GAGAGACTATATAAACCCAGTCTTGTGTTTAAAAACCTGTTCTTTCAA 4395
Qy 1084 AlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGlu 1103
Db 4396 TGTTTAGAAGGGGCAAACTGCAGATTTGACTTAAAGGTGTGTGTAGACCTTATGCCAGAG 4455
Qy 1104 LeuThrTrpLeuLeuAsnGlyGlnProValProAspAlaSerHisLysMetLeuVal 1123
Db 4456 ACGTCTCGTTTCATGATGGCCACCAAAATGTCAATGACTATACCATAAAGTAGTCATT 4515
Qy 1124 ArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThr 1143
Db 4516 AAAGAAGATGTTACTCAATCACTAATTAATTTGCTCCCTGCCACACCCAGTGAATCTGGGAA 4575
Qy 1144 TyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuLeuLeuSerVal 1163
Db 4576 TGGACTGTGGTGGCCAAAACAGGGCAGAGACTCTCAATTTTCAGTCAATTTTAACTGTG 4635
Qy 1164 ValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyVal 1183
Db 4636 GAAGCTGGGAACATCAGGTAAACCCGATGTTGTAGAAAACTGAAAAATGTCAATATA 4695
Qy 1184 ProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProProValPhe 1203
Db 4696 AAGGAAGTTTCCCGACTTGAATGAAAGTCAGAGCTACGGGTAAACCCCAACCTGACATT 4755
Qy 1204 TyrTrpLysLysAspAsnGluThrIle---ProCysThrArgGluArgIleSerMetHis 1222
Db 4756 GTATGGTTGAAAAACAGTGACATCATTTGCGCTCATAAATATCCAAAATCAGAAATT--- 4812
Qy 1223 GlnAspThrThrGlyTyrAlaCysLeuIleGlnProAlaLysLysSerSerAlaGly 1242
Db 4813 GAAGGAACCAAGGAGAGAGCTGCCCTTAAATCCATGCCACTGTCCAGCAAGATTCGCC 4872
Qy 1243 TrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 1262

```
||||| 4873 TGTATACGCGACTGCTATTATAAAGCTGGCAGACACACTACAAGATGCAAGTAAAT 4932
||||| 1263 IletyrAlaGlntrpHisGlnIleProProMetSerValArgProSerGlySer 1282
||||| 4933 GTTGAAGTTGAGTTGCAGAGCCTGAGCCAGAGAGAAAGTTAATCATCCACCGGGGACA 4992

RESULT 11
AAS05390
ID AAS05390 standard; DNA; 81940 BP.
XX AAS05390;
XX 24-OCT-2001 (first entry)
XX Human titin (connectin) gene sequence.
XX Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
XX titin-related disease; zebrafish; heart failure; heart disease; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 133..80913
XX /*tag= a
XX /product= "titin"
XX WO200151666-A1.
XX 19-JUL-2001.
XX 12-JAN-2001; 2001WO-US01212.
XX 12-JAN-2000; 2000US-0175787.
XX (GEO ) GEN HOSPITAL CORP.
XX Fishman MC;
XX WPI: 2001-451869/48.
XX P-PSDB; AAU05396.
XX Determining if a subject has or is at risk of developing a
XX titin-related disease or condition, particularly heart failures,
XX comprises detecting the presence of a mutation in the titin gene
XX -
XX Disclosure; Page 35-57; 114pp; English.
XX The present sequence encoding for human titin (also known as connectin)
XX is described in an invention relating to a novel method for determining
XX whether a subject has or is at risk of developing a titin-related
XX disease or condition. The method comprises analysing a nucleic acid of
XX a sample from the subject and detecting the presence of a mutation
XX (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
XX titin gene, which indicates that the subject has or is at risk of
XX developing a titin-related disease. The zebrafish which has a phenotype
XX similar to mammalian heart failure is used as a model. The method is
XX useful for detecting an increased likelihood of heart disease, such as
XX heart failure, in a patient, so that appropriate intervention can be
XX instituted before any symptoms occur. The method may also be used to
XX facilitate determination of etiology of an existing heart condition,
XX such as heart failure, to identify compounds that can be used to treat
XX or prevent heart conditions, in prenatal genetic screening, e.g. to
XX identify parents who may be carriers of a recessive titin mutation.
XX Compounds identified using the methods may be used to treat patients
XX that have or are at risk of developing heart disease, e.g. heart
XX failure.
XX Sequence 81940 BP; 26373 A; 17100 C; 18899 G; 19568 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3,78e-30 Length: 81940
XX Score: 806.50 Matches: 359
```

```
Percent Similarity: 34.15% Conservative: 229
Best Local Similarity: 20.85% Mismatches: 512
Query Match: 11.71% Indels: 622
DB: 22 Gaps: 54

US-09-818-990B-2 (1-1320) x AAS05390 (1-81940)

QY 100 LeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGln 119
||||| 82 CTTAGTCCCACCTTTTAGGCACTCTTGAGAAATCAGAGTG----- 123
DB 82 CTTAGTCCCACCTTTTAGGCACTCTTGAGAAATCAGAGTG----- 123
QY 120 AspAsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGln 139
||||| 124 -----CCTAGAAAGATGACAAC-----CAAGCACCGACG 153
DB 124 -----CCTAGAAAGATGACAAC-----CAAGCACCGACG 153
QY 140 TyrCysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGlu 159
||||| 154 TTTACGAGCGGTTACAAAGCGTTGTGTA-----CTGGAG 189
DB 154 TTTACGAGCGGTTACAAAGCGTTGTGTA-----CTGGAG 189
QY 160 GluLeuSerSerPheLysSerHisSerLysArgIleArgProArgAlaCysLys 179
||||| 190 GGTAGTACCGCAACCTTTGAGGCTCATTAGTGGTTTCCAGTCTCTGAGTGAGCTGG 249
DB 190 GGTAGTACCGCAACCTTTGAGGCTCATTAGTGGTTTCCAGTCTCTGAGTGAGCTGG 249
QY 180 AsnHisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPhe 199
||||| 250 TTTAGGGATGGCCAGGTGATTTC-----ACTTCACCTCTGCCGCGTGCAGATCTCCTTT 306
DB 250 TTTAGGGATGGCCAGGTGATTTC-----ACTTCACCTCTGCCGCGTGCAGATCTCCTTT 306
QY 200 SerAspLeuSerGluArgArgGluArgSerSerValProIleProAlaAspThr 219
||||| 307 AGCGAT-----GGCCGCGCTAAACTGACGATCCCGCGGTGACT 345
DB 307 AGCGAT-----GGCCGCGCTAAACTGACGATCCCGCGGTGACT 345
QY 220 ArgAspAsnGluValAsnHisAlaLeuGluGlnGluAlaLysArgGluAlaGlu 239
||||| 346 AAGCCACACAGTGGACGATATTCCTG----- 372
DB 346 AAGCCACACAGTGGACGATATTCCTG----- 372
QY 240 GlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerLeu 259
||||| 373 -----AAAGCCCAATGGATCTGGACAACGAGCTAGTACTGCTGAGCTTCTC 420
DB 373 -----AAAGCCCAATGGATCTGGACAACGAGCTAGTACTGCTGAGCTTCTC 420
QY 260 TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGlu 279
||||| 421 GTCAAGAGCTGAGACA-----GCACCCACCACTTCCTTCAACGACTGCAGAGATGACC 474
DB 421 GTCAAGAGCTGAGACA-----GCACCCACCACTTCCTTCAACGACTGCAGAGATGACC 474
QY 280 ValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGln 299
||||| 475 GTGAGACAAGCAAGCAAGTGAGACTCCAAGTGAGAGTGACTGGAATCCCTACACCTGTG 534
DB 475 GTGAGACAAGCAAGCAAGTGAGACTCCAAGTGAGAGTGACTGGAATCCCTACACCTGTG 534
QY 300 ValArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGln 319
||||| 535 GTGAAGTTCTACCGGGATGGAGCGAAATCCAGAGCTCCCTTGATTTCCAAATTTCAAA 594
DB 535 GTGAAGTTCTACCGGGATGGAGCGAAATCCAGAGCTCCCTTGATTTCCAAATTTCAAA 594
QY 320 AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr 339
||||| 595 GAAGGAGCACTTACAGCTTACTGATTGCAAGAGCATACCTGAGGACTCAGGAGACCTAT 654
DB 595 GAAGGAGCACTTACAGCTTACTGATTGCAAGAGCATACCTGAGGACTCAGGAGACCTAT 654
QY 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
||||| 655 TCAGTAATGCCACCAATACGTTGGAGAGAGCTACTTTCGACTGCTGAATTACTGCTTCAA 714
DB 655 TCAGTAATGCCACCAATACGTTGGAGAGAGCTACTTTCGACTGCTGAATTACTGCTTCAA 714
QY 360 Gly-----ValSerSerAspSerGluGlyAspProAsn 371
||||| 715 GGTGAAGAAGATGACCTGCTTAAAGACAAAGACAAATTTTCGACTGCTCAGATCTCA 774
DB 715 GGTGAAGAAGATGACCTGCTTAAAGACAAAGACAAATTTTCGACTGCTCAGATCTCA 774
QY 372 LysGluGluMetAsnArgIleGlnLys----- 380
||||| 775 GAATCAGACAAACCCGAAATGAAAGAGAGATTGAGCCCACTTTGTGATGCCAGATCAATT 834
DB 775 GAATCAGACAAACCCGAAATGAAAGAGAGATTGAGCCCACTTTGTGATGCCAGATCAATT 834
QY 381 -----Pro 381
||||| 835 GCAACAGTTGAGATGGTCATAGATGGTCCGCTGGGCAACAGCTGCCACATATAAACACCT 894
DB 835 GCAACAGTTGAGATGGTCATAGATGGTCCGCTGGGCAACAGCTGCCACATATAAACACCT 894
QY 382 AsnGluValSerProProThrThrSerAlaValIleProProAlaValProGlnAla 401
||||| 382 AsnGluValSerProProThrThrSerAlaValIleProProAlaValProGlnAla 401
```

Db 895 CCCAGGATTCTCCGAAGCCMAAGTCAAGATCCCCCAACACACCACCGCTCTATTGCTGCCAAA 954
Qy 402 GlnHisLeuValaIaGlnProArgValaIaThrIleGlnGlnCysGlnSerProThrAsn 421
Db 955 GCACAGCTGGCTCGCAGCAGCTCCCATGCCCATAGACACTCCCTCTCCCGGTGAGA 1014
Qy 422 TyrLeuGlnGlyLeuAspGlyLysProIle----- 431
Db 1015 CACGTGGGGCAGCAGCCCATCTCCCGTCAAGTCCGTGCTCCAGCAGCAAGAATCTCC 1074
Qy 432 -----IleAlaIaProValaIaPheThrLysMetLeuGlnAsnLeuSer 445
Db 1075 ACATCCCCATFCAGGTCTGTAGTCTCCATGTCTATGCGTAAGACTCAGGCATCCACC 1134
Qy 446 AlaSerGluGlyGlnLeuValaIaPheGluCysArgValLysGlyAlaProSerProLys 465
Db 1135 GTGCCACAGGCTCTGAAGTG-----CCTCCCCCT--- 1164
Qy 466 ValGluTrpTyrArgGluGly-----Thr 473
Db 1165 -----TGAAGCAAGAGGGCTACGTGGCTCTCATCTGAGGCTGAGATGAGAGAGACA 1218
Qy 474 LeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGlu 493
Db 1219 ACGCTGACAACTCTACTCAGATCAGACAGACAGAGAGATGGGAAGGAGATACGGGTGC 1278
Qy 494 ProGluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 1279 CAGGAGCAAGTACCATCAGTGGTCTGCGGGTCTGCCGCCAGTGTCTGGCCAGTGCT 1338
Qy 514 ThrCysThrAla-----SerAsnLys 520
Db 1339 AGCTACGCAGCAGAGCGTGTGCCACTGGTGCTAAAGAGGTGAACAAGATGCTGCACAAA 1398
Qy 521 TyrGlyThrValSerSerIle----- 527
Db 1399 AGTCAGCTGTGGGACTGTTGTGCGCTTGATATGCCAGAGTGAGAGAACCAAGTG 1458
Qy 528 -----AlaGlnLeuHisValArgGly 534
Db 1459 ATCAGCGCTGTAGACAGACTGCTCAGAGGACACACGACTGCTGTGCGCATCCACCT 1518
Qy 535 AsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAla 554
Db 1519 GCTCAAGAACAGGTAAAGAAAGGAGCGGAGAGACTGCTGTAAGTAGTAGTGCC 1578
Qy 555 Ala-IleGluProGlnProSerProProHisSerGluPro----- 567
Db 1579 GCCGTAAGCCAGGAACAGAAATTAATCAAGAACCAAGAAATAATTACCACAAAG 1638
Qy 567 ----- 567
Db 1639 CAAGAGCAGATGCACGTAACATCATGACAGATAGAAGAAAGAACTGAAACAAATTGTGA 1698
Qy 568 -----ProSerValGluGlnProProLysProLys-----Le 578
Db 1699 CCAAAGGTAGTAATTTCCGCGAGCTAAAG-----CCAAAGAAACAAGAAACTAGAAATTTCT 1752
Qy 578 uGluGlyValLeuValAsnHisAsnGlu----- 587
Db 1753 GAAGAAATTACTAAGAAACAAGAACTAAGTAAGTAAGAAACAATAATGAAGGAACATAGG 1812
Qy 588 -----ProArgSerSerArgIle-GlyL 596
Db 1813 AAAACAGTTGTACTAAAGTCATAGTTGCCACACCCCAAGTCAAGAAACAAGATTTAGTA 1872
Qy 596 eu-----ArgValHisPhe-----AsnLeuProGlu 605
Db 1873 TCAAGAGGTAGAGAGGCATTACTACCAAAAGAGAACAAAGTGCAAAATAACTCAGGAGAAG 1932
Qy 605 spAspLysGlySer-Glu-----Ala 611
Db 1933 ATGAGAAAGGAGCCGAGAAAACTGCTTGTCTACAATAGCAGTTGCTACTGCTAAAGCC 1992

Qy 612 SerSerGluAlaGlyValValThrThrArgGlnThr----- 623
Db 1993 AAAGAAACAAGAAACAATACTACTAGAGAACTATAGAGAACTATGCTACTAGACAAGAAATC 2052
Qy 623 ----- 623
Db 2053 CAAGTTACCATGAAAGGTGGAGCTTGGAAAAAGGCTGAAGCTGTAGCAACAGTTGTT 2112
Qy 624 -----ArgProAspSer***GlnGluArg 631
Db 2113 GCTGCAGTAGACACAGGCCCGAGTCAGAGAGCCAGAGAGCTTGGGCATCTTGAAGAAATC 2172
Qy 632 PheAsnGlyGlnAlaThr-----LysThr 639
Db 2173 TATGCTCAGCAGACCACTTTGGAGTACGGATATAAAGAACGCATCTTCCGCCGCAAGGTA 2232
Qy 640 ProGlu---ProSerPheProValLysGluPro-----Pro 650
Db 2233 GCTGAGCCTCCCAACGTCAGGCTCAGAACCCACACGTTGCCCTAAAGCAGTCAAGCCT 2292
Qy 651 ProValLeuAlaLysProLys-----LeuAspSerThrGlnLeuGlnGlnLeuHis 667
Db 2293 AGAGTAATCCAGGCTCCTCTGTGAGACTCATATAAACTACTGATCAAAAGGGAATGCAC 2352
Qy 668 AsnGlnValLeuLeuGluGln-----HisGlnLeu 677
Db 2353 ATATCATCAGATCAAGAAACTACAGATCTAACACGGAAAGATTAGTCCATGTGGAT 2412
Qy 678 GlnAsnProProProSerProLysGluPheProPhe***MetThrValLeuAsn--- 696
Db 2413 AAAGCCCGCCGACAGCTAGCCCTCAC-----TTTACTGTTCAAAAAATTTCTGTT 2463
Qy 697 -----SerAsnAlaProAlaValThrThrSer*** 707
Db 2464 CCTAAGACAGAACATGGATATGAGCATCAATACCGGCTAGTGTCTATGTCACATTACAA 2523
Qy 708 LysGlnValLysAlaProSerSerGln-----ThrPheSerLeuAlaArgProLys 724
Db 2524 AAAGAGTTGTGAGCCACATCTTCTGCTCAGAGATCACCAATCGGTGAAGGCTCCTACT 2583
Qy 725 TyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSerSerProVal 744
Db 2584 GTG-----AAGCCAGTAGACTAGAGTAGAGGAGGAGCCACACCCCTTGGCCACAG 2634
Qy 745 PheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeu 764
Db 2635 TTCCCTTCGCTGCACACACCATACTTACAAG----- 2667
Qy 765 ValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluPro 784
Db 2668 ---AGTGAAGCTGGCGTTGAGTGAAGGAGAGTGGGTGAGCATCACTGGCACCACC 2724
Qy 785 Leu----- 785
Db 2725 GTCCGTGAAGAGCGCTTTGAAGTACTGCAGCGCGCAAGCCAGAGTAACAGAAACAGCA 2784
Qy 786 ---ProProGlyProThrGlu---ProThrProProPheThrPheSer----- 800
Db 2785 AGAGTACCAGCACTGTTGAATTCCTGTACTCCACCAACTTTGGTCTCGGGCTTAAAA 2844
Qy 801 -----IleProSerGlyAsnGlnPheGlnProArgCys---ValSerProIlePro 816
Db 2845 AATGTGACTGTCATAGAGGTGAATCTGTCACTTGGAGTGGCCACATCTCTGGNATACCCA 2904
Qy 817 ValSerProThr-----SerArgIleGlnAsn 825
Db 2905 ---TCCCGGAGTGACATGGTACAGGGAAGACTTACCAAAATCGAAAGTTCATTGACTTC 2961
Qy 826 ProValAlaPheLeuSerSerVal----- 833
Db 2962 CAGATAACCTTCCAGAGTGAATTCGCTTATGATTCGGAAGCAATTTGGGGAAGAC 3021

```
QY 833 ----- 833
Db 3022 AGCGGCGATTACTTGCAGTGTGTAAATGAGCGTGAACCGTFCAGCACATCCTGCTAT 3081
QY 833 ----- 833
Db 3082 CTGGCTGTGCAGGTGTGAGAAGAAATTTGAAAAGGAAACACACAGCCGTGACTGAGAAATTT 3141
QY 834 -----
Db 3142 ACTACAGAAGAAACCCCTTTGTGTAGCTCAAGAGATGTGTTATGACTGATAGTACGCCCTC 3201
QY 847 ***LeuProArgSerAlaProSerMetProSer----- 857
Db 3202 ACAGAGAAACAGCAGCGCTGGAGAACCTGCCGCGCTTACTTTATTACAAAACACAGTG 3261
QY 858 ---GlnGlyLeuAlaLysLysAsnThr----- 868
Db 3262 GTCAGAAACTGTTGGAAGGTGGAGCGTGGTGTGGATGCCAAGTTGGCGGCAACCCA 3321
QY 869 GlnPro----- 870
Db 3322 AAGCCCATGTATACTGAAAAAATCTGGTGTCTCTAACCACTGGATACAGATACAAA 3391
QY 870 ----- 870
Db 3382 GTGAGTTACAACAACAAACCGGTGAATGCAAGCTGGTATTCTATGACTTTTGTCTGAT 3441
QY 870 ----- 870
Db 3442 GATGCTGGAGAATACACTATTGTTGTCGAATAAGCATGAGAAACTTCTGCATCTGCT 3501
QY 870 ----- 870
Db 3502 TCCTTGCTTGAAGAGCTGATTATGAGCTTACTGATGAAGTCCCGAGCAAGAAATCCTTAT 3561
QY 871 ----- 879
Db 3562 CAGACACAAGTGACTGCATTGTTCAAGAACCTGAAGTTGGAGAAACAGCACCTGGATT 3621
QY 880 -----
Db 3622 GTATFACTCTGAGTATGAAAAAGAGTATGAAAAGAAAGCAAGCCTTAATTAGGAAGAAATG 3681
QY 888 GlyLysLysIleThrPheSerAspValArgProAsnGlnGluTyrLysIleSerSer 907
Db 3682 GCCAAACTACTGTAGTGGTCAGAACTATGTAGAAATCAGGAATTCATATTTCTTCC 3741
QY 908 PheGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrPro----- 924
Db 3742 TTTCAAGAGAGACTATTAAAGAAATTTAAGATATAGATATAATAAGACTACATTAGAGAA 3801
QY 925 ---ValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCysIle 943
Db 3802 CTTCTTGAAGAAGATGGAGAGAAAGATGGCAGCTTGACATTTCTGAATCTGGAAGCTGT 3861
QY 944 AlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProValThr 963
Db 3862 GAATCAGATTTGATTTAAGATTCAGAAATATAGAAATTTCTGAGGGGATGGGTGCTACT 3921
QY 964 PheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGlyLys 983
Db 3922 TTTTATTGCAAGATGCTGGATATCCATTACCAAGATTTGCTTGTACAAAGATGGCAAG 3981
QY 984 GlnIleSerLysArgAsnGluHisCysLysMetArgGluGlyAspGlyThrCysSer 1003
Db 3982 CGCATC---AAACATGGAGAAAGATACCAATGGACTTCTTACAAGATGGCAGAGCTAGT 4038
QY 1004 LeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsn 1023
Db 4039 CTGGTATACCTGTGTCTTCCAGAGATGAAGAAATCTACACTGCATTTGCCAGCAAT 4098
QY 1024 ProGlnGlyArgIleSerCysSerGlyHisLeuMetValGln----- 1037
```

```
Db 4099 ATTAAGGAAATGCAATTTGCTCAGGAAATTTGATGTGGAGCCTGCTGCACCACCTTGA 4158
QY 1038 ----- 1039
Db 4159 GCTCCGACTTACATTCACACACTAGACCCAGTGAAGCAATCAGATCTCTCTCTCCACGT 4218
QY 1040 -----ProIleArg-----SerArgLeuThr 1046
Db 4219 TCAGTGAGCAGGTCTCCTATACGATGTCTCCTCAGCGATGTCCCTGCAAGGATGTCT 4278
QY 1047 SerAlaGlyGlnSerHisArgGlyArgSer-----ArgValGlnGluArgAspLys 1063
Db 4279 CCTGCAGGATGTCCCTGCAAGAAATGTCCCTGACGTAGCTGGAGGACAGATGAG 4338
QY 1064 GluProLeuGlnArgPheArgPheArgPheLeuGlnAlaProGlyAspMetVal 1083
Db 4339 TCACAACTT---GAGAGACTATATAAACCACTTTGTTGTTAAACCTGTTCTTCAAA 4395
QY 1084 AlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGlu 1103
Db 4396 TGTTTGAAGGGCAAACTGCCAGATTTGACTTAAAGGTGTTGGTAGACCTATGCCAGAG 4455
QY 1104 LeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuVal 1123
Db 4456 ACGTTCTGTTTCATGATGCCAGCAAAATGTCAATGACTATACCCATAAAGTAGTCAT 4515
QY 1124 ArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThr 1143
Db 4516 AAGAAGATGTTACTCAATCACTAATATTGTCCTGCCACCCAGCTGATTCTGGGGAA 4575
QY 1144 TyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuLeuLeuSerVal 1163
Db 4576 TGGACTGTGTTGCCAAACACAGGGCAGGAGATCTCAATTTTCAGTGATTTTAACTGTG 4635
QY 1164 ValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyVal 1183
Db 4636 GAAGCTGTGGAACATCAGGTAAAAACCGATGTTGTAGAAAAAATGAAAAATGTCAATATA 4695
QY 1184 ProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProProValPhe 1203
Db 4696 AAGGAAGTTTCCCGACTGGAATGAAAGTCAGAGCTACGGGTAAACCCCAACCCAGCAT 4755
QY 1204 TyrTrpLysLysAspAsnGluThrIle---ProCysThrArgGluArgIleSerMetHis 1222
Db 4756 GTATGTTGAAAAACAGTGACATCATTTGCTCTATAAATATCCAAATCAGAAAT--- 4812
QY 1223 GlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGly 1242
Db 4813 GAAGGAACCAAGGAGAGAGCTGCCCTTAAATTCGATTCCACTGTCAGCAAGATTTCTGCC 4872
QY 1243 TrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 1262
Db 4873 TGGTATATCTGGCACTGTATTATAAAGCTGGCAGACACATACAAGATGCAAGATTAAT 4932
QY 1263 IleTyrAlaGlnTrpHisGlnIleProProMetSerValArgProSerGlySer 1282
Db 4933 GTTGAAGTTGATTTGCAGAGCCCTGAGCCAGAGAAAGTTAATCATCCACGGGGGACA 4992

RESULT 12
AXA40026
ID AXA40026 standard; DNA; 790 BP.
XX
AC AXA40026;
XX
DT 02-JUL-1999 (first entry)
XX
DE Prostate cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
```


PS Disclosure; Page 23-25; 43pp; English.

XX The present sequence encodes a human myotilin (myofibrillar protein
CC with titin-like Ig-domains) protein. The myotilin protein is a component
CC of the striated and cardiac muscle cytoskeleton. The protein contains two
CC C2-type immunoglobulin (Ig)-like domains with homology to certain
CC Ig-like domains of titin. Myotilin resides both in the sarcomere, where
CC it localises within the I-bands and is bound to alpha-actinin, and along
CC the sarcolemmal membrane. The myotilin gene locates in chromosome 5q31
CC inside a 2 Mb region, which contains the LGMD1A disease gene.
CC Transfection of myotilin into mammalian cells induces formation of
CC thick actin bundles. The myotilin protein is used for preparing compounds
CC that regulate actin-cytoskeleton and cell growth, and cancer and
CC microbial infections. The protein is also used in the preparation of
CC medicines for treating cancer and antimicrobial infection.

XX Sequence 2244 BP; 737 A; 506 C; 416 G; 585 T; 0 other;

Alignment Scores:

Pred. No.: 1,62e-29 Length: 2244
Score: 760.00 Matches: 220
Percent Similarity: 45.95% Conservative: 98
Best Local Similarity: 31.79% Mismatches: 220
Query Match: 11.03% Indels: 155
DB: 21 Gaps: 19

US-09-818-990B-2 (1-1320) x AAZ61262 (1-2244)

QY 653 LeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeu 672
DB 98 CTACCAAGCAGGAGCAGCAGTAGTATCTCAGGATCTCAACAAGGAGAGCAGCAAGGTT 157
QY 673 GluGlnHisGlnLeuGlnAsnProProSerSerSerProLysGluPheProPhe**Met 692
DB 158 GCTTCTGATTCCTTACAA-----CCTCGGTAAATCCAGGCTGTGGCC----- 202
QY 693 ThrValLeuAsnSerAsnAlaProAlaValThrThrSer***LysGlnValLysAla 712
DB 203 -----CAAATTCAGGCCCCACCCTCCAGGAACAATCATTTATAGTAATAATTG 253
QY 713 ProSerSerGlnThrPheSerLeuAla-----ArgProLysTyrPhePhe 727
DB 254 CCTTCATCTTCCATATACCACTAAGCATGTTAACTACGAACTGCCAAACACTTCATC 313
QY 728 ProSerThrAsnThrAlaAlaThrValAlaProSerSerSerProValPheThrLeu 747
DB 314 CAGTCCAAACCCATGCTGGCTCAGATTCGAGCTCCTCGACCAGAAACCTCCAGCTTC 373
QY 748 SerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuVal-SerHi 767
DB 374 TCTAGC-----CAGACCAACAG-----TCTTCCATATATCATCCAGCCC 412
QY 767 sProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProPr 787
DB 413 CGCCAGGTACAGAGCAAA-----GATTTCTCGCTCC 445
QY 787 o-----GlyProThrGluProThrProProProPheThrPheSerIleProSerGl 804
DB 446 TCAACACTGAGCTCTCAATACACCATCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 505
QY 804 YAsn-GlnPheGlnProArgCysValSerProIleProValSer-ProThrSerArgile 823
DB 506 CATGCTGGCTCCACCCAGG-----CCAAAGGTTACAAACCACTATATAC 550
QY 824 GlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIleProProThr 843
DB 551 CAGTCCCGCAGCAGCTCTCCAGCTCAGCTCATATATACCATCACAG-----CCTGATTAC 601
QY 844 AsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLys 863
DB 602 AATGACAGTAAATTCCTTCGCTTGGATTCCAACTATCAACAGCTCCTCAGCTGGCCAA 661
QY 864 -----AsnThrLysSerProGlnProValAsn-----Asp 873

DB 662 CCTATAAATGCAAGCCATCCCAAAATGCAAAATGCTAAGCCCATACCAAGAACTCCTGAT 721
QY 874 AspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPhe 893
DB 722 CATGAATAACAGATCAAAAGAGAGCTTTGATTCAAGATTTGGAAGAAAGCTGAATGC 781
QY 894 SerAspValArgProAsnGlnGlnGluTyrLysIleSerPheGluGlnArgLeuMet 913
DB 782 AAGACACC-----CTTCTTCAATATGGAATCAACCTCTCAACA 820
QY 914 AsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGln 933
DB 821 TATGAAGA-GAA----- 831
QY 934 HisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHis 953
DB 831 ----- 831
QY 954 PheArgValThrGluGlySerProValThrPheThrCysLysIleValGlyIleProVal 973
DB 831 ----- 831
QY 974 ProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
DB 832 -----GATGGCTCGCAGATTCTAGGACCACAGAA----- 861
QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSer-AspAs 1013
DB 862 -----TGCAGCTGCTGTGTTTCAAGC-----TCAGGATGA 891
QY 1013 pAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHi 1033
DB 892 CAGTGGT-----GCACAGACTCCGACACACACCACTCAGAACATGCGCG 936
QY 1033 sLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisAr 1053
DB 937 ACTGCAAGTTCTTACATACACAAAGTAAGAGTAGATCAACCTCAAGGGAGATGTGAAT-- 994
QY 1053 gGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgPr 1073
DB 995 -----GATCAGGATGCAATCCAGGAGAAATTTTCCACC 1029
QY 1073 oHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAs 1093
DB 1030 ACGTTTCATTCAAGTCCAGAGACACATGTCGATTGATGAGGAAGATCTCCACAANTGGA 1089
QY 1093 pCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVa 1113
DB 1090 CTTCAAAGTGAGTGGCTGCGAGCTCCTGATGTCATGTTCTATAAATGGAAGAACAGT 1149
QY 1113 lLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIl 1133
DB 1150 TCAATCAGATGATTGCACAAAATGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1209
QY 1133 eAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGl 1153
DB 1210 TGAAGTAGTCAGACTTCAGATGAGGGGCTTATGTCATGTTGCCAAGATATAGAGCAGG 1269
QY 1153 YGlnAsnSerPheSerLeuGluLeuSerValAlaLysGluValLysLysAlaProVa 1173
DB 1270 AGAAGCCACCTTCACTGTCAGCTGGATGTCCTTGCACAAAGAACATATAAGAGCACAAT 1329
QY 1173 lLeuLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCy 1193
DB 1330 GTTTATCTACAAACACACAGAGCAAAAAGTTTATAGGGAGATTCAGTGAAACTAGATG 1389
QY 1193 sArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGluThrIlePr 1213
DB 1390 CCAGATCTCGGCTATACCTCCCAACAGCTTCTCTGAAAAAGAAATATGAATGGGTACA 1449
QY 1213 oCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIl 1233

```
Db 1450 ATTCAACACTGACCGAATAAGCTTATATCAAGATAACACTGGAAGAGTTACTTTACTGAT 1509
Qy 1233 eGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaG1 1253
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 1510 AAAGATGTAACAAGAAGATGCTGGGTGTATCTGTGTACAGCTTATATGAAGCTGG 1569
Qy 1253 yIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnIleProPr 1273
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 1570 AGTGACTACATGTAAACAAGATTAGCGTTACGCGACGTCCAAACCAAACTCTCCAGC 1629
Qy 1273 oPro-----MetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysG1 1291
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 1630 TCCTAAGCAGTTACGGGTTCCAGCAACATTCCAGCAAAATATTATACACTTAATGGGAAGG 1689
Qy 1291 yLeuAspIlePheSerAlaPheSerMetMetGluSerThrMetValTyrSerCysSerSe 1311
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 1690 TTTGAATGTAAACAAGCTTTTAAACCA--GAAGGAGAATTTCCAGCGTTTGGCAGCTCA 1746
Qy 1311 rArgSerValValGluSerAspGluLeu 1320
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 1747 ATCTGGACTCTATGAAGTGAAGAACTT 1774

RESULT 14
AAx40027/c
ID AAx40027 standard; DNA; 747 BP.
XX AC
XX AAX40027;
XX
XX
XX
XX 02-JUL-1999 (first entry)
XX
XX Prostate cancer associated gene.
XX
XX Cancer associated antigen; diagnosis: research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
XX Homo sapiens.
XX
XX W09904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0102322.
XX
XX 17-JUL-1997; 97US-0896164.
XX
XX 10-OCT-1997; 97US-0061599.
XX
XX 10-OCT-1997; 97US-0061765.
XX
XX 10-OCT-1997; 97US-0948705.
XX
XX 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI; 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Claim 67; Page 639-640; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
```

```
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
XX SQ Sequence 747 BP; 156 A; 180 C; 192 G; 218 T; 1 other;
XX
Alignment Scores:
Pred. No.: 1.93e-25 Length: 747
Score: 668.50 Matches: 130
Percent Similarity: 75.81% Conservative: 33
Best Local Similarity: 60.47% Mismatches: 48
Query Match: 9.70% Indels: 4
DB: 2.20 Gaps: 2
US-09-818-990B-2 (1-1320) x AAX40027 (1-747)
Qy 1109 AsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArg-GluThrGlyVa 1128
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 745 GATGAAAGCGGTACGCTTGACAGGCTCCCAAGATGCTGGTGCCTAAAGAACGGGGT 686
Qy 1128 lHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAl 1148
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 685 GCACCTCTGTATCATAGAGCCAGTCACGTACGTGATGCCGGCATCTACACATGTATAGC 626
Qy 1148 aThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluVa 1168
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 625 TACCACCGCAGCAGGAGAGAACTCATTACGCTGGAGCTTGTGGTCTGCTGAAGAAGC 566
Qy 1168 lLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisPr 1188
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 565 ACACAAACCCCTGTGTTATTGAGAAGCTCCAAAACACAGGAGTTGCTGATGGGTACCC 506
Qy 1188 oValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAs 1208
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 505 AGTGGCGCTGGAATGCTGTATTGGGAGTGCCACCACTCAGATATTTTGGAAAGAAGA 446
Qy 1208 pAsnGluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTy 1228
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 445 AATGAATCACTACTCCTACAGCAGCTGACCGAGTGAGGATGCCACGAGCAACACCGCTA 386
Qy 1228 rAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAl 1248
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 385 CATCTGCTGCTCATTTCAGGAGGCCACAAAAGAAGATGCTGGGTGATATACTGTGTCAGC 326
Qy 1248 aLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHi 1268
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 325 CAAGATGAAGCAGGAGATTGTGCTCTACTGCCAGGCTGGACGCTTTACACCCAGTGSCA 266
Qy 1268 sHis-----GlnIleProProMetSerValArgProSerGlySerArgTyrGlySe 1286
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 265 TCACGAGTCACAGCAGCACCAAGCAAAAAGTACGGCCCTCCACCGAGTCGCTATGAGC 206
Qy 1286 rLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerSerMetGluSerThrMetVa 1306
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 205 ACTTCGGACCAGGAGCTAGACATCAAGCAGCGCTTCCACCT---GAGGCCAACCCATC 149
Qy 1306 lTyrSerCysSerArgSerValValGluSerAspGluLeu 1320
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 148 TCACCTGACACTGAATACTGCTTGGTAGAAAGTAGGAGACCTG 106
RESULT 15
AAC75084
ID AAC75084 standard; cDNA; 392 BP.
XX AC
XX AAC75084;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF639 polynucleotide sequence SEQ ID NO:1277.
XX
XX
```

KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW	vulnery; antipsoriatic; antiparkinsonian; neurotic; neuroprotective;
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW	antiviral; antibacterial; antifungal; antineumatic; antithroid;
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive; ss.

OS Homo sapiens.

XX PN WO200058473-A2.

XX PD 05-OCT-2000.

XX . 31-MAR-2000: PE

XX 31-MAR-1999: PR

PR 02-APR-1999;
PR 05-APR-1999;

PR 30-MAR-2000;
XX

PA (CURA-) CURA
XXPI Shimkets RA,
XX

DR WPI; 2000-60
DR P-PSDB: AAB4

XX Novel nuclei

PT	useful for t
PT	neurodegener

XX PS Claim 5: pag

XX AAC74446 to CC

CC which represents sequences having

CC antipsoriatic;
CC osteopathic;

CC	immunostimul
CC	antidiabetic

CC	antiinflamm
CC	antithyroid;

CC the presence
CC pathological

CC nucleic acid
CC vectors. The

CC proliferativ
CC graft vs hos

CC hypertension
CC erythematous

CC bacterial or
CC allergies, a

CC nocturnal ha
CC coagulation;

XX
SQ
Sequence 392

Alignment Scores:

Pred. No.:	Score:
------------	--------

Percent Similarity
Best Local Similarity

Query Match:
DB:

US-09-818-990B-2 (1-1320) x AAC75084 (1-392)	
Qy	806 GlnPheGlnProArgCysValSerProIleProValSerProThrSerArgIleGlnAsn 825
Db	1 CAGTTTCAGGCCCGCGCTGTGTCCCAATTCCTGTCTCTCTACCAAGCGGATTTCAGAAC 60
Qy	826 ProValAlaPheLeuSerSerValLeuProSerLeuProAlaIleProProThrAsnAla 845
Db	61 CCAGTGGCTTTCCTCAGCTCTGTCTGCCCTCTCTCTCTGCCATCCACCACCAATGCC 120
Qy	846 Met***LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThr 865
Db	121 ATGGGGCTGGCTAGAAAGTGACCATCCATGCCATCCAGGAGATTAGCGAGAAAAATACA 180
Qy	866 LysSerProGlnProValAsnAspAspAsnIleArgGluThrLysAsnAlaValIleArg 885
Db	181 AAGTCTCCTCAACAGTGAAATGATGATAACATTCGTGAACCTAAGAACGCGAGTTCGA 240
Qy	886 AspLeuGlyLysLysIleThrPheSerAspValArgProAsnGlnGlnIuThrLysIle 905
Db	241 GACTTGGGAAAAAATAACTTTTCAGTGTGTCTAGAGCAACCAACAGCAGGAGTAGTACAAAAT 300
Qy	906 SerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProVal 925
Db	301 TCAAGCTTTGAGCAGAGGCTGATGAATGAATAAGAGTTTCGCTTGGAACGCTACTCTCGTT 360
Qy	926 AspGluSerAspAspGluIleGlnHisAsp 935
Db	361 GATCAATTCACATGATGAATTCACATGAT 390

Search completed: November 30, 2002, 19:20:52
Job time : 574 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 30, 2002, 19:10:45 ; Search time 76 Seconds
(without alignments)
5326.491 Million cell updates/sec

Title: US-09-818-990B-2
Perfect score: 6890
Sequence: 1 MDDSIESTATISQILLRESY.....MESTVMYSCSSRVVESDEL 1320

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/uspto.spool/US09818990/runat_26112002_093405_23250/app.query.fasta_1.1479
-DB-Issued Patents_NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1.1.37 @runat_26112002_093405_23250 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5b_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	760	11.0	2244	US-09-354-151-1	Sequence 1, Appli
2	327	4.7	2793	US-08-795-868-13	Sequence 13, Appl
3	327	4.7	2793	US-09-303-069-13	Sequence 13, Appl
4	327	4.7	2793	US-09-134-250-13	Sequence 15, Appl
5	326	4.7	2614	US-08-795-868-15	Sequence 15, Appl
6	326	4.7	2614	US-09-303-069-15	Sequence 15, Appl
7	326	4.7	2614	US-09-134-250-15	Sequence 15, Appl
8	300	4.4	4608	US-09-041-886-24	Sequence 24, Appl
9	300	4.4	4608	PCR-US94-05277-1	Sequence 1, Appli
10	278.5	4.0	6814	US-09-484-970B-66	Sequence 66, Appl
11	278	4.0	4092	US-07-757-022B-51	Sequence 51, Appl
12	275	4.0	3420	US-07-757-022B-103	Sequence 103, App

13	275	4.0	4215	4	US-07-757-022B-61	Sequence 61, Appli
14	275	4.0	5008	4	US-07-757-022B-1	Sequence 1, Appli
15	272	3.9	3874	4	US-09-877-730-31	Sequence 31, Appl
16	266.5	3.9	3963	4	US-07-757-022B-45	Sequence 45, Appl
17	266.5	3.9	3963	4	US-07-757-022B-59	Sequence 59, Appl
18	264.5	3.8	3943	4	US-08-506-296B-27	Sequence 27, Appl
19	263.5	3.8	4086	4	US-07-757-022B-39	Sequence 39, Appl
20	258	3.7	4843	3	US-08-986-485-1	Sequence 1, Appli
21	257	3.7	3453	4	US-09-877-730-7	Sequence 7, Appli
22	257	3.7	49377	1	US-08-764-233A-1	Sequence 1, Appli
23	256	3.7	1143	4	US-09-877-730-3	Sequence 3, Appli
24	256	3.7	2715	4	US-09-877-730-5	Sequence 5, Appli
25	256	3.7	2958	4	US-09-877-730-9	Sequence 9, Appli
26	256	3.7	3210	4	US-09-877-730-1	Sequence 1, Appli
27	255	3.7	1493	2	US-08-752-307B-6	Sequence 6, Appli
28	255	3.7	1493	4	US-09-707-802-6	Sequence 6, Appli
29	255	3.7	1493	4	US-09-991-326-6	Sequence 6, Appli
30	250.5	3.6	28958	1	US-08-258-261B-6	Sequence 6, Appli
31	250.5	3.6	28958	1	US-08-456-837-6	Sequence 6, Appli
32	250.5	3.6	28958	1	US-08-457-342-6	Sequence 6, Appli
33	250.5	3.6	28958	1	US-08-457-646A-6	Sequence 6, Appli
34	250.5	3.6	28958	1	US-08-458-076A-6	Sequence 6, Appli
35	250.5	3.6	28958	1	US-08-764-233A-4	Sequence 4, Appli
36	250.5	3.6	28958	1	US-08-457-335A-6	Sequence 6, Appli
37	250.5	3.6	28958	1	US-08-729-214-6	Sequence 6, Appli
38	250.5	3.6	28958	3	US-09-028-934-6	Sequence 6, Appli
39	247	3.6	4279	4	US-09-041-886-22	Sequence 22, Appl
40	246	3.6	3219	4	US-09-877-730-17	Sequence 17, Appl
41	245	3.6	3360	1	US-08-408-093-5	Sequence 5, Appli
42	245	3.6	3360	1	US-08-408-420A-5	Sequence 5, Appli
43	245	3.6	3360	1	US-08-714-901-5	Sequence 5, Appli
44	245	3.6	3360	3	US-08-040-741-5	Sequence 5, Appli
45	245	3.6	3945	4	US-07-757-022B-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1
US-09-354-151-1
; Sequence 1, Application US/09354151
; Patent No. 6458929
; GENERAL INFORMATION:
; APPLICANT: CARPEN, Olli
; APPLICANT: GRONHOLM, Mikaela
; APPLICANT: HEISKA, Leena
; APPLICANT: MYKANEN, Olli-Matti
; APPLICANT: SALMIKANGAS, Paula
; TITLE OF INVENTION: Myotilin, A No. 6458929a1 Actin-Organizing Protein
; FILE REFERENCE: 0933-0142P
; CURRENT APPLICATION NUMBER: US/09/354,151
; CURRENT FILING DATE: 1999-07-16
; EARLIER APPLICATION NUMBER: 60/093,169
; EARLIER FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2244
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (281)..(1774)
US-09-354-151-1

Alignment Scores:
Pred. No.: 1.98e-46
Score: 760.00
Percent Similarity: 45.95%
Best Local Similarity: 31.79%
Query Match: 11.03%
Indels: 155
Gaps: 19
Length: 2244
Matches: 220
Conservative: 98
Mismatches: 220

US-09-818-990B-2 (1-1320) x US-09-354-151-1 (1-2244)

Qy	653	LeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeu	672
		: : :	
Db	98	CTACCAAGCCAGCAGCACAGTATTCTCAGGATCTCAACAAGCAAGCAGACCAAGAGTT	157
Qy	673	GluGlnHisGlnLeuGlnAsnProProSerSerProLysGluPheProPhe***Met	692
Db	158	GCTTCTGATTCCTTTACAA-----CCTTCGGTAATTCAGGCGTTGGGCC-----	202
Qy	693	ThrValLeuAsnSerAsnAlaProAlaValThrThrSer***LysGlnValLysAla	712
Db	203	-----CAAATTCAGGCCCCACCCCTTCCAGGACAACATCATATTAGTAATAATTG	253
Qy	713	ProSerSerGlnThrPheSerLeuAla-----ArgProLysTyrPhePhe	727
		: : : : : : :	
Db	254	CCTTCATCTTCATATACCAACTAAGCATGTTTAACATACGAAGCTCCAAACACTTCATC	313
Qy	728	ProSerThrAsnThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeu	747
		: : : : : : :	
Db	314	CAGTCCCAAAACCATGTGGCTCCAGATTCGACGCTCCCTGGACCAACCTCCAGCTTC	373
Qy	748	SerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeuVal-SerHis	767
		: : : : :	
Db	374	TCTAGC-----CAGACCAACAG-----TCITCCATTCATCCAGCCC	412
Qy	767	sProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuPro	787
		: : : : : : :	
Db	413	CGCCAGTGTACAGACAAA-----GATTTCTGCCTCC	445
Qy	787	O-----GlyProThrGluProThrProProPheThrPheSerIleProSerG	804
		: : : : :	
Db	446	TCACACTGAGCTCTCAATCACCATGCTCTCTCTGCTTCCTGCTCTCCCCACGAG	505
Qy	804	yAsn-GlnPheGlnProArgCysValSerProIleProValSer-ProThrSerArgIle	823
		: : : : : : :	
Db	506	CATGCTGCTCCCAACCCAG-----CCAAAGGGTTACAACCCACTATAAC	550
Qy	824	GlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIleProThr	843
		: : : : : : :	
Db	551	CAGTCCCCAGCAGCTCTCTCAGCTCCATATACCATCAG-----CCTGATTAC	601
Qy	844	AsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLys	863
		: : : : : : :	
Db	602	AATGACGATAAAATCCCTTCGCTATGGATCCAACTATCAACAGTCCCTCAGCTGCCAA	661
Qy	864	-----AsnThrLysSerProGlnProValAsn-----Asp	873
		: : : : :	
Db	662	CCTATAATGCAAGCCATCCCAACTGCAAAATGCTAAGCCCATCAAGAACTCTGTAT	721
Qy	874	AspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysIleThrPhe	893
		: : : : : : :	
Db	722	CATGAATACAAAGGATCAAAAGAAAGCTTTGATTCAAGATTTGAAAGAAGCTGAAATGC	781
Qy	894	SerAspValArgProAsnGlnGlnIleLysIleSerSerPheGluGlnArgLeuMet	913
		: : : : : : :	
Db	782	AAGGACACC-----CTTCTTCATAATGGAATCAACGCTCTAACA	820
Qy	914	AsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGln	933
		: : : : :	
Db	821	TATGAAGA-GAA-----	831
Qy	934	HisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuHis	953
		: : : : : : :	
Db	831	-----	831
Qy	954	PheArgValThrGluGlySerProValThrPheThrCysLysIleValGlyIleProVal	973
		: : : : : : :	
Db	831	-----	831
Qy	974	ProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLys	993
		: : : : : : :	
Db	832	-----GATGGCTCCGACGATTCAGTTCAGGACACAGAA-----	861

APPLICANT: Lee, Mu-En
APPLICANT: Hsieh, Chung-Ming
TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
TITLE OF INVENTION: AND STRIATED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,868
FILING DATE: 06-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,577
FILING DATE: 22-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/032001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 3...1983
OTHER INFORMATION:
US-08-795-868-13

Alignment Scores:
Pred. No.: 1,5e-14 Length: 2793
Score: 327.00 Matches: 165
Percent Similarity: 35.35% Conservative: 63
Best Local Similarity: 25.58% Mismatches: 254
Query Match: 4.75% Indels: 165
DB: 2 Gaps: 26
US-09-818-990B-2 (1-1320) x US-08-795-868-13 (1-2793)
QY 35 AlaGluProSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGly 54
Db 409 AGCAGCCAGTCGAGCGCGGCCACCCGTCGGGACCCCC-GGGGC----- 455
QY 55 GlyGlyGlnAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGlu 74
Db 456 -----TCGCAGGAAGAACTCGGGCG 476
QY 75 SerValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThr 94
Db 477 CCAGGCAGCGTGGCGGCGCGCGCGCTGTTCAG-----CAGAAAGCGGCTCGCTG 530
QY 95 GlnAlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGlu 114
Db 531 GACGAGCGACGCT-----CAGCGCGCCCGGCTCAGACCTCGAG 572
QY 115 ProAsnPheCysGlnAsp-----AsnProArgSerProThrSerSerLysGlu----- 130
Db 573 CTGGCTTCGCCAGGAGCTGGCGCGCTCCAGCTCCAGCTCGGGAGGAGCTGGT 632

QY 131 -----SerProGlnGluAlaLys 136
Db 633 CGTCGCACAGTCCCTCGCGGCCACACCTCGAGCGTCCCATCCCTCGAGAGCCCGGC 692
QY 137 ArgProGlnTyrCysSerGluThrGlnSerLysVal-----PheLeuAsnLysAla 154
Db 693 GAGCCCGCGCTTCTCTCGGCCCTCCACCCCAAGACATCGCGGGCGGTGAGCCCGCC 752
QY 155 AlaAspPheIleGluGluLeuSerSerLeuPheLysSerHisSerSerLysArgIleArg 174
Db 753 GCGCCCGCAGCGCCCTCTCCGAGCAGCGCGGAGAAGCG--GGGACGAGCCTGGAGG 809
QY 175 ProArgAlaCysLysAsnHisLysSerLysLeuGlu-----SerGlnAsnLys 190
Db 810 CCCAGGAGC---CGCGCGCGCGCGGAGCAGACAGCGCGGGGGAAGCCCGCAGAG 866
QY 191 ValMetGlnGluAsnSerSerPheSerAspLeuSerGluArgArgGlu---ArgSer 209
Db 867 GTTAGCGTCGGGACCAATTCCCGCTGACCCGCGAGCAGCCATCCAGGAGTGCAGGAGC 926
QY 210 ServAlProIleProIle-----ProAlaAspThrArgAspAsn----- 222
Db 927 CCTGTGCGCGCCCGCGCGCGATCCCGCAGAGCGCAGAGAAACACCCCGCGTGG 986
QY 222 ----- 222
Db 987 AAGCGGAGCCCGCGCGCAGCGTTCCTGCGCTCGGCGCACCGCGGGCGCTGGAG 1046
QY 223 -----GluValAsnHisAlaLeuGlnGlnGlnAlaLys 234
Db 1047 GCGCTGCTGTACCCAGACACTTGGAGAAGAACAGCGGGCGGCTGAGGACAGAGAAGG 1106
QY 235 ---ArgArgGluAlaGluGln-----AlaAla 242
Db 1107 CTTGCGAGAGCGCGGAGGAGCGTCCCTGGGGCGCTGGGACCGCGGCGGCGCGC 1166
QY 243 SerGluAlaAlaGlyGlyAspThrThrProGlySer-----SerProSerSer 258
Db 1167 AGCAGGGCAAGGTGCGCGGGCGCGCCACCTCCCTGAGCTCGAGTCTTCGGATGAC 1226
QY 259 LeuTyrTyr-----GluGluProLeuGlyGlnProPhePheThrGlnLysLeu 275
Db 1227 TCCTACGTGTCGCTGGGAGAGAGCCCTA--GAGGCGCCCTGTGTGAGATCCCTG 1283
QY 276 ArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValGlyIle 295
Db 1284 CAGAAATGTTGTTGGCACCAGGGCGCAGATGTGCTGCTCAAAATGTATCATCATCTCCAAC 1343
QY 296 ProProGlnValArgTyrTyrCysGluGlyLysGluLeuGluAsnSerProAspIle 315
Db 1344 CCGCGCGCCCAAGTCTCTGGCACAAGGTGGTTCAGCGCTGCGCAGCGAGGCGCGCTC 1403
QY 316 HisIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluAsp 335
Db 1404 CTCCTCGGCTGAGGTGAGCGGCACACCTCTGCTCAGGAGGCGGCGGACGAGAT 1463
QY 336 ThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGlu 355
Db 1464 GCGGGAGCTATATGGCCACCGCACCAACGAGTGGGCCAGCCACCTGTGCGCGCTCA 1523
QY 356 IleTyrIleGluGlyValSerSerSerAspSerGluGlyAspProAsnLysGluGluMet 375
Db 1524 CTGACCGTGAGACCGGTGGGTCTACATCCCTTTTCAGCAGCCCATCATCCTCGAGAG 1583
QY 376 AsnArgIleGlnLysProAsnGluValSerSerProThrThrThrSerAlaValIlePro 395
Db 1584 GAATACCTGAGCCCGCCAGAGGAGTTCACAGAGCTGGGAGACC----- 1628
QY 396 ProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGln 415
Db 1629 -----TGCGCGCGCAACCCCGCCCATCAAGGCC 1655

```
Qy 416 CysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleAlaAlaPro 435
Db 1656 AGTCCAGCCAGAACCGCGTCTTCTGACACTGGCTCAAAG-----GCACCCCCC 1706
Qy 436 ValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGlu 455
Db 1707 ACCTTCAAGGTCTCACTTATGACACAGCTCAGTAAGAGAGGCCAAGATGTCATCATGAGC 1766
Qy 456 CysArgValLysGlyAlaProSerProLysValGluTyrArgGluGlyThrLeuIle 475
Db 1767 ATCCCGGTGAGGGAGGCCACGAGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1826
Qy 476 GluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGlu 495
Db 1827 CGC-----CCAGAC-----CAGCGCGCTTTGCGGAGGAGGCT 1859
Qy 496 Glu-----IleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 1860 GAGGTGGGCTGTGCGGCTCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1919
Qy 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
Db 1920 ACTTGCAAGCGGTCAAGTAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1979
Qy 534 GlyAsnGluAspLeuSerAsnAsnGlySerLeuHis----- 545
Db 1980 GGC-----GAGTG-AGCTCAGGGGGGCACCTGCGCTCCCGCTACCTCCGAGCGCGC 2032
Qy 546 -----SerAlaAsnSerThrThrAsnLeu-----AlaAlaIleGluProGln--- 559
Db 2033 GCCCTGTCTCAGGCACCTCTCGGACCTCGCTGTGTTTCACTGCTCTGCTGCTGCTGCTGCTG 2092
Qy 560 -----ProSerProHisSerGluProProSer 569
Db 2093 CAGTCGCGCGCGGACCGCTCCAGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2152
Qy 570 ValGluGlnProPro 574
Db 2153 ATAGCCCATGGGCC 2167
```

RESULT 3

```
US-09-303-069-13
; Sequence 13, Application US/09303069A
; Patent No. 6350592
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; FILE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
; FILE REFERENCE: 05433/039001
; CURRENT APPLICATION NUMBER: US/09/303.069A
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: US 09/134,250
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 2793
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1985)
US-09-303-069-13
```

Alignment Scores:

Pred. No.:	1.5e-14	Length:	2793
Score:	327.00	Matches:	165
Percent Similarity:	35.35%	Conservative:	63
Best Local Similarity:	25.58%	Mismatches:	254
Query Match:	4.75%	Indels:	165
DB:	4	Gaps:	26

```
US-09-818-990b-2 (1-1320) x US-09-303-069-13 (1-2793)
Qy 35 AlaGluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 54
Db 409 AGCAGCCCAAGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455
Qy 55 GlyGlyGlnAspAspLeuProAspLeuSerAlaPheLeuSerGlnGluGluLeuAspGlu 74
Db 456 -----TCGCAGGAAGAAGTCTCGGGCG 476
Qy 75 SerValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThr 94
Db 477 CCAGCGCAGCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 530
Qy 95 GlnAlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGlu 114
Db 531 GACGAGCCACGCGT-----CAGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 572
Qy 115 ProAsnPheCysGlnAsp-----AsnProArgSerProThrSerSerLysGlu----- 130
Db 573 CTGCGCTTCGCGCAGGAGCTGGCGCGCATCCGCGCTCCACGTCGCGGGAGAGCTGGTG 632
Qy 131 -----SerProGlnGlnAlaLys 136
Db 633 CGCTCGCAGAGTCCCTGCGCGCCACGCTGCAGCGTGCCTCCCTCCCTCCGAGAGCCCGCG 692
Qy 137 ArgProGlnTyrCysSerGluThrGlnSerLysVal-----PheLeuAsnLysAla 154
Db 693 GAGCCCCGCTCTCTCTGCGCCCTCCACCCCAAGACATCGCGGGCGGTGAGCCCGCGCG 752
Qy 155 AlaAspPheIleGluGluLeuSerSerLeuPheLysSerHisSerSerLysArgIleArg 174
Db 753 GCCGCCAGCGCGCTCTCCGAGCAGCGCGGAGGCGG-----GGGAGCAGCCTGGGAGG 809
Qy 175 ProArgAlaCysLysAsnHisLysSerLysLeuGlu-----SerGlnAsnLys 190
Db 810 CCCAGGAGC---CGCGGGCGCGCGGAGGAGCAGACAGCGGGGAGGCGCGCGCGCGCGAG 866
Qy 191 ValMetGlnGluAsnSerSerPheSerAspLeuSerGluArgArgGlu---ArgSer 209
Db 867 GTTAGGGCTCGGGACCAATTCCTGCTGACCCGGAGCAGAGCCATCCAGGAGTGCAGGAGC 926
Qy 210 SerValProIleProIle-----ProAlaAspThrArgAspAsn----- 222
Db 927 CCTGTGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 986
Qy 222 ----- 222
Db 987 AAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1046
Qy 223 -----GluValAsnHisAlaLeuGluGlnGlnGluAlaLys 234
Db 1047 GCGCGTGTGTACCCACAGACCTTGGAGAGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106
Qy 235 ---ArgArgGluAlaGluGln-----AlaAla 242
Db 1107 CTTCGAGAGGCGCGGAGGAGGAGCGTCCCTGGGGCGCTTCGAGCGCGCGCGCGCGCGCG 1166
Qy 243 SerGluAlaAlaGlyGlyAspThrProGlySer-----SerProSerSer 258
Db 1167 AGCCAGGSCAAAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1226
Qy 259 LeuTyrTyr-----GluGluProLeuGlyGlnProProArgPheThrGlnLysLeu 275
Db 1227 TCCTACGTGCTCGCTGGAGAAGAGCCCTCA---GAGGCGCGCTGTGTTTGAGATCCCGCTG 1283
Qy 276 ArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValIleGlyIle 295
Db 1284 CAGAATGTGGTGGTGGCCAGGAGGAGAGTGTGCTGCTCAAAATGTATCATCACTACCTCAAC 1343
Qy 296 ProProProGlnValArgTyrCysGluGlyLysGluLeuGluAsnSerProAspIle 315
Db 1344 CCCCCGCCCCAAGTGTCTTGGCACAAGGATGGGTCAAGCGCTGCGCAGCGAGGCGCGCGCTC 1403
```



```
QY 316 HistleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAsp 335
   :::      |||      |||      |||      |||      |||      |||      |||
Db 1404 CTCCTCGGGCTGAGGTGAGCGGCACACCTGCTGCTCAGGGAGGCCAGGCAGCAT 1463
QY 336 ThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGlu 355
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1464 GCCGGAGCTATATGGCCACCGCACCAAGAGCTGGGCCAGGCCACCTGTGGCGCTCA 1523
QY 356 IleTyrIleGluGlyValSerSerSerSerSerSerSerSerSerSerSerSerSerSer 375
   :::      :::      :::      :::      :::      :::      :::      :::
Db 1524 CTGACCTGAGACCGGTGGTCTACATCCCTTTTCAGCAGCCCATCATCTCGAGCAG 1583
QY 376 AsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAlaValIlePro 395
   :::      |||      |||      |||      |||      |||      |||      |||
Db 1584 GAATACCTGAGCCCCCAGAGGAGTCTCCAGAGCCTGGGAGACC----- 1628
QY 396 ProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGln 415
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1629 -----TGGCCGCGAACCCCAACCATCATGAAGCCC 1655
QY 416 CysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaPro 435
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1656 AGTCCCAAGCCAGAACCCCGCTTCTTGACACTGGCTTCCAAG-----GCACCCCCC 1706
QY 436 ValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGlu 455
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1707 ACCTTCAAGTCTCACTATGGACCACTAGTAGAGAGCCCAAGATGTCATCATGAGC 1766
QY 456 CysArgValLysGlyAlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIle 475
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1767 ATCCGCTGAGGGGAGCCCAAGCCTGTGCTCTCTGCTGAGAAACCCGACGCCCGCTG 1826
QY 476 GluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGlu 495
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1827 CGC-----CCAGAC-----CAGCGCGGCTTTTGGGAGGAGGCT 1859
QY 496 Glu-----IleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1860 GAGGTGGGCTGTCCGGCTCGCATCTCGCTCAGAGCGTGGCGATGCTGTTCTTCTAC 1919
QY 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1920 ACTTGCAAGCGGTCAATAGTAGTGTGCTCGCAGTGCAGGCGCCGCTTGGAGGTCCGA 1979
QY 534 GlyAsnGluAspLeuSerAsnAsnGlySerLeuHis----- 545
   |||      |||      |||      |||      |||      |||      |||      |||
Db 1980 GGC-----GAGTG-AGCTCAGGGGGCCACTGCGCTCCCGCCGCTACCTTCGAGCCGC 2032
QY 546 -----SerAlaAsnSerThrThrAsnLeu-----AlaAlaIleGluProGln--- 559
   |||      |||      |||      |||      |||      |||      |||      |||
Db 2033 GCCCTGTCTCAGGCACCTCTCGGACCTCGCTGTGTTTCACTGCTCTGCCACAGACC 2092
QY 560 -----ProSerProProHisSerGluProProSer 569
   |||      |||      |||      |||      |||      |||      |||      |||
Db 2093 CAGTGGCGGCGCGGACCGCTCCAGAGCTCCCTTCCCAACCCATCGAGCCCGCCAGGGG 2152
QY 570 ValGluGlnProPro 574
   |||      |||      |||      |||      |||      |||      |||      |||
Db 2153 ATAGCCCATGGGCC 2167

RESULT 4
US-09-134-250-13
; Sequence 13, Application US/09134250B
; Patent No. 6393753
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; FILE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
; FILE REFERENCE: 05433/038001
; CURRENT APPLICATION NUMBER: US/09/134,250B
; CURRENT FILING DATE: 1998-08-14
```

```
; EARLIER APPLICATION NUMBER: US 08/795,868
; EARLIER FILING DATE: 1997-02-06
; EARLIER APPLICATION NUMBER: US 08/494,577
; EARLIER FILING DATE: 1995-06-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 2793
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1985)
US-09-134-250-13

Alignment Scores:      1.5e-14      Length:      2793
Pred. No.:      327.00      Matches:      165
Score:      35.35%      Conservative:      63
Percent Similarity:      25.58%      Mismatches:      254
Best Local Similarity:      4.75%      Indels:      165
Query Match:      4      Gaps:      26
DB:

US-09-818-990b-2 (1-1320) x US-09-134-250-13 (1-2793)

QY 35 AlaGluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 54
   :::      |||      |||      |||      |||      |||      |||      |||
Db 409 AGCAGCCCAAGTCGGAGCGCGGCCACCGCTGGGGCACCCCC-GGGGCC----- 455
QY 55 GlyGlyGlnAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGlu 74
   |||      |||      |||      |||      |||      |||      |||      |||
Db 456 -----TCGCAGGAAGAACTCGCGGCG 476
QY 75 SerValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThr 94
   |||      |||      |||      |||      |||      |||      |||      |||
Db 477 CAGGCGAGCTGGCGGAGCGCGCGCTGTTCCAG-----CAGAAAGCGGCTCGCTG 530
QY 95 GlnAlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGlu 114
   |||      |||      |||      |||      |||      |||      |||      |||
Db 531 GACGAGCGCAGCGT-----CAGCGCAGCCCGGCTCAGACTCGAG 572
QY 115 ProAsnPheCysGlnAsp-----AsnProArgSerProThrSerSerLysGlu----- 130
   |||      |||      |||      |||      |||      |||      |||      |||
Db 573 CTGCGCTTCGCCCAAGAGTGGCGCGCATCGCGCTCCACGTCGCGGAGGAGCTGGTG 632
QY 131 -----SerProGlnGluAlaLys 136
   |||      |||      |||      |||      |||      |||      |||      |||
Db 633 CGCTCGCAGCAGTCCCTCGCGCCACGCTGCAGCGTGCCTCCCTCGAGAGCCCGCGC 692
QY 137 ArgProGlnTyrCysSerGluThrGlnSerLysLysVal-----PheLeuAsnLysAla 154
   |||      |||      |||      |||      |||      |||      |||      |||
Db 693 GAGCCCGCGCTCTCTCTCGGCCCTCCACCCCAAGACATCGCGGCGCTGAGGCCCGCC 752
QY 155 AlaAspPheIleGluLeuSerSerLeuPheLysSerHisSerSerLysArgIleArg 174
   |||      |||      |||      |||      |||      |||      |||      |||
Db 753 GCCGCCCAGCGCCCTCTCCGAGCAGCGCGGAGAACGCG---GGGGACGACCTGGGAGG 809
QY 175 ProArgAlaCysLysAsnHisLysSerLysLeuGlu-----SerGlnAsnLys 190
   |||      |||      |||      |||      |||      |||      |||      |||
Db 810 CCCAGGAGC---CGCGGGCGCGCGGAGGACAGACGCGGGGGAAGGCCCCAGCAGAG 866
QY 191 ValMetGlnAsnSerSerSerPheSerAspLeuSerGluArgArgGlu---ArgSer 209
   |||      |||      |||      |||      |||      |||      |||      |||
Db 867 GTTAGCGGTGGGACCAATTCCTGACCCGCGGAGCAGACCATCCAGAGTGCAGGAGC 926
QY 210 SerValProIleProIle-----ProAlaAspThrArgAspAsn----- 222
   |||      |||      |||      |||      |||      |||      |||      |||
Db 927 CCTGTGCGGCCCGCGCGCGATCCCCCAGAGGCCAGGACGAAAGACACCCCGCTGG 986
QY 222 ----- 222
   |||      |||      |||      |||      |||      |||      |||      |||
Db 987 AAGCGGGAGCCCGCGGCGCAGGCGGTGCGCTTCTTCCTGGCCACGCGCGGCGCTGGAG 1046
```

```
Qy 223 -----GluValAsnHisAlaLeuGluGlnGlnAlaLys 234
Db 1047 GCGCTGCTGTACCCAGACCTTGGAGAAAGACAGGGCGGCGCTGAGGCAGAGAGG 1106
Qy 235 ---ArgArgGluAlaGluGln-----AlaAla 242
Db 1107 CTTCGCAGAGCGCGGAGAGACGGTCCCTGGGGGCCCTGGGACCGCGGAGGGCGCGC 1166
Qy 243 SerGluAlaAlaGlyAspThrProGlySer-----SerProSerSer 258
Db 1167 AGCCAGGCAAGAGTGGCGGGCGCGGCCACCTCCCTGAGCTCGAGTCTTCGGATGAC 1226
Qy 259 LeuTyrTyr-----GluGluProLeuGlyGlnProArgPheThrGlnLysLeu 275
Db 1227 TCCTAGGTGTCCGCTGGAGAGAGCCCTA---GAGGCCCTGTGTGAGATCCCGCTG 1283
Qy 276 ArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIle 295
Db 1284 CAGAATGTGTGGTGGCACCGAGGCGAGATGTCTCAATGTATCATCATCATGCCAAC 1343
Qy 296 ProProGlnValArgTyrCysGluGlyLysGluLeuGluAsnSerProAspIle 315
Db 1344 CCCCCGCCCAAGTGTCTGGCACAGATGGGTACGGCTGCCAGCGGAGGGCGGCTC 1403
Qy 316 HisIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAsp 335
Db 1404 CTCCTCGGGGTGAGGTGAGCGGCACACCTCTGCTCAGGGAGCGGCAGGCAGCAGAT 1463
Qy 336 ThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGlu 355
Db 1464 GCGGGAGCTATATGGCCACCGCCACCAACAGAGTGGGCCAGGCACCTGTGCGCGCTCA 1523
Qy 356 IleTyrIleGluGlyValSerSerAspSerGluGlyAspProAsnLysGluGluMet 375
Db 1524 CTGACCGTGACCGCGTGGTGTCTACATCCCTTTCAGCAGCCGCCATCACCCTCCGACGAG 1583
Qy 376 AsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAlaValIlePro 395
Db 1584 GAATACCTGAGCCCCCAGAGAGAGTTCAGAGAGCTGGGAGACC-----1628
Qy 396 ProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGln 415
Db 1629 -----TGGCGCGGAACCCACCATGAAGCCC 1655
Qy 416 CysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaPro 435
Db 1656 AGTCCAGCCAGACCGCGGTCTCTGACACTGGCTCCAG-----GCACCCGCC 1706
Qy 436 ValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGlu 455
Db 1707 ACCTTCAAGGTCTCACTTATGGCAGCTCAGTAAGAGAGGCCAAGATGTTCATCATGAGC 1766
Qy 456 CysArgValLysGlyAlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIle 475
Db 1767 ATCCCGGTGACGGGGAGCCCAAGCCTGTGCTCTCCGTGGTGAGAAACCGCCAGCCCGGTG 1826
Qy 476 GluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGlu 495
Db 1827 CGC-----CCAGAC-----CAGCGCGCTTTCGGAGGAGGCT 1859
Qy 496 Glu-----IleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 1860 GAGGTGGGTGTGCGGCTCGGATCCTGGCTGCAGAGCGTGGCGATGCTGTGTTTCTAC 1919
Qy 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
Db 1920 ACTTGCAAGCGGTCATAGATATGGTGTCTCGGCAGTGCAGAGCGCGCTGTGGAGGTCCGA 1979
Qy 534 GlyAsnGluAspLeuSerAsnAspGlySerLeuHis-----545
Db 1980 GGC-----GAGTG-AGCTAGGGGGCCACCTGCTGCCCTCCCGCTACCTCCGAGCGCG 2032
Qy 546 -----SerAlaAsnSerThrThrAsnLeu-----AlaAlaIleGluProGln--- 559
```

```
Db 2033 GCCCTGTCTCAGGACACCTCTCGGACCTCGTGTGTTTCACTGCTCTGCCACAGACC 2092
Qy 560 -----ProSerProHisSerGluProProSer 569
Db 2093 CAGTGTGGCGCCGAGACCCGTCGCCAGCTCCCTCCCAACCCCATGACGCCCCAGGGGG 2152
Qy 570 ValGluGlnProPro 574
Db 2153 ATAGCCCATGGGCC 2167

RESULT 5
US-08-795-868-15
; Sequence 15, Application US/08795868
; Patent No. 5846773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
; TITLE OF INVENTION: AND STRIATED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...1803
; OTHER INFORMATION:
;
US-08-795-868-15

Alignment Scores:
Pred. No.: 1.6e-14 Length: 2614
Score: 326.00 Matches: 166
Percent Similarity: 35.10% Conservative: 79
Best Local Similarity: 23.78% Mismatches: 294
Query Match: 4.73% Indels: 160
DB: 2 Gaps: 25

US-09-818-990B-2 (1-1320) x US-08-795-868-15 (1-2614)
Qy 6 IleGluAlaSerThrSerIleSerGlnLeuArg---GluSerTyrLeuAlaGluThr 24
Db 421 ATCCGCCGATCTACGTGCGGGAGGAGCTGTCGCTTCGCACGAGATCCCTCGTGCACAC 480
```

```
QY 25 ArgHisArgGlyAsnAsnGluArgSerArgAlaGluPro----- 37
D 481 CTGAGGGCGCCCATCCCTCGGAGCGCGGAGGCGCCCACTCTCTCCCGGCTTCC 540
QY 38 -----SerSerAsnProCysHisPheGlySer 46
D 541 ACACCCAGACCTCACGGGTGTGAGCGCGGTCCACCGCCG-----CGCCCT 591
QY 47 ProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeuProAspLeuSerAlaPhe 66
D 592 CTTAGTGTGGGGGAAA---TCTGGGACGAGCTGGGAGGCGCCGAAACAGAGGCGCG 648
QY 67 LeuSerGlnGluLeuAspGluSerValAsnLeuAlaAlaGluLeuAlaLeuAsnTyrAsp 86
D 649 GTGGCAGGACTGAACCGGGGGAAGCGCCGACAGGAGATCAAGCTCGGACCAATTC 708
QY 87 ProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeuSerProAspGlnMetLys 106
D 709 CCGCTAACACAGAGAGAGCCATCCAGAGTGCAGG-----744
QY 107 HisSerProAsnLeuSerPheGluProAsnPheCysGlnAspAsnProArgSerProThr 126
D 745 ---AGCCCT-----GTGCGCCCTACACCGCGGATCCCCGGAGAGCAGGACA 789
QY 127 SerSer-----LysGluSerProGlnGluAlaLysArgProGlnTyrCys 141
D 790 AAGCCCTCCCGTCTCGCAAGCGGACCCCTCTCAAGCGGTGCGC-----837
QY 142 SerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheLeuGluLeu 161
D 838 -----TTTCTGCGTGGCCCACTCCGGGAGTGGAGGAC---870
QY 162 SerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsnHis 181
D 871 -----TCTGTTCTGCCCCCAAACTTGGAGAGAAT 900
QY 182 LysSerLysLeuGluSerGlnAsnLysValMetGln-----GluAsnSerSerPhe 199
D 901 AGAGCGGACCCGAGGTGAGAGAGCTTCGCAGAGACCTTGAGGAGGATGCCCTTGG 960
QY 200 SerAspLeuSerGluArgGluArgSerSerValProIleProAlaAspThr 219
D 961 GGGCCCTGGGACCGCAGAGGAGCCCGCAGC-----990
QY 220 ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnAlaLysArgArgGluAlaGlu 239
D 991 -----CAAGGCAAAAGTCCCGGTCTCGG 1014
QY 240 GlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeu 259
D 1015 CCTACTTCCCCGAGCTCGAGTCTCAGACGACTCTATGTGCTCGGTGG-----1065
QY 260 TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGlu 279
D 1066 -----GAAGAGGCCCTG---GAGGCACCGGTGTTTGTAGATCCCTCTGCAGAAATATGGT 1116
QY 280 ValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGln 299
D 1117 GTGGCGCAGAGGTGACGGTGTACTTAAGTGTATCATCCGCAACCCCGCCACCCCAA 1176
QY 300 ValArgTyrPyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGln 319
D 1177 GTGTCTGGAAAAGATGGGTCCATGTTGCACAGCAGGCGTCTCTTCATCCGGGCT 1236
QY 320 AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr 339
D 1237 GAAGGTGAACGGCACACACTGCTCTCAGAGAGGCCCGGAGTGTGTGTTGGAGGTAC 1296
QY 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
D 1297 ACAGCCACTGCCCAACCAAGTGGGCCAAGTACCTGTCTTCTTCACTGGGTGTGAGA 1356
```

```
QY 360 GlyValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGln 379
D 1357 CTTGGGCGCTCCACATCCCTTTTCAGCAGCCCCCATCATCTGTATGAGGAGTACCTGAGC 1416
QY 380 LysProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValPro 399
D 1417 CCCCCAGAGAGTTCAGAGCCTGGGGAGAC-----1449
QY 400 GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerPro 419
D 1450 -----TGGCCCCGAAACCCCTACCATGAGCTCAGTCCAGCCAGCAG 1488
QY 420 ThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLys 439
D 1489 GATCATGATTCTCCGACTCTTCTTCCAAG-----GCACCCCAACGTTCAAGGTC 1539
QY 440 MetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLys 459
D 1540 TCACCTCATGACCAATCGGTGAGAGAAGTCAAGATCTCATTTATGAGCATCCCGTGTGAG 1599
QY 460 GlyAlaProSerProLysValGluTyrTyrArgGluGlyThrLeuIleGluAspSerPro 479
D 1600 GGAGAGCCCAAGCTGTGGTTCCTGCTGAGGAATCGACAGCCCGTGGC-----CCA 1653
QY 480 AspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGlu-----Ile 497
D 1654 GAC-----CAGCGGCGCTTTCAGAGGAGGCGGAGGTGGGCTC 1692
QY 498 CysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAla 517
D 1693 TGCGCTGTAGGATCTCGGTGCTGTAACGGGGCGATCTGCTTCTACATCATCAAGCGC 1752
QY 518 SerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAsp 537
D 1753 GTCAACAAATATGGCGTCGCGAGTCGCGAGCCCGCTGGAGGTCCGAGGCGAG-----1806
QY 538 LeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGlu 557
D 1807 ---TGAGCTCAGGGGGCCACCTCGGCTGCCCGCTACCTCCGAGCTGCACCCCTGTCT 1863
QY 558 ProGlnPro-----SerProHisSerGluPro-----567
D 1864 CAGCAGCTCTCGGACCTCGCTGTGTTTCTACTGCTCTGCCACACAGCCCGGCTC 1923
QY 568 -----ProSerValGluGlnProProLysProLysLeuGluGlyValLeuValAsnHis 585
D 1924 CCGCGCCGAGCTTAGCCCATGTCCCTTCCCTCCCTCCTCCTCCTCCTCCTCCTCCT 1968
QY 586 AsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsnLeuProGluAsp 605
D 1969 ACAGCACCC-----TGGGGTAACCCACCGCGGC 1995
QY 606 AspLysGlySerGluAlaSer-SerGluAlaGlyValValThrThrArgGlnThrArgPr 625
D 1996 CCCTGTGTGATCC-----TCCCTCCCAAGTGGATATGTGCTGTGCAGACAGAGGCC 2049
QY 625 oAspSer**GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu---ProSerPh 644
D 2050 CCCAGAGG-ACTGAGTGTGGGAAGGAGGCGCATGAGGGGTGCCAAGCTCCCTCGGTC 2108
QY 644 eProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
D 2109 TCCCATAGGAGGAGCATCCAGCGAGTGTGCTATGCTGCTGTACAGGCCACT 2160
```

RESULT 6

```
US-09-303-069-15
; Sequence 15, Application US/09303069A
; Patent No. 6350592
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
```

```
; FILE REFERENCE: 05433/039001
; CURRENT APPLICATION NUMBER: US/09/303,069A
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: US 09/134,250
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1806)
US-09-303-069-15

Alignment Scores:
Pred. No.: 1,6e-14 Length: 2614
Score: 326.00 Matches: 166
Percent Similarity: 35.10% Conservative: 79
Best Local Similarity: 23.78% Mismatches: 294
Query Match: 4.73% Indels: 160
DB: 4 Gaps: 25

US-09-818-990b-2 (1-1320) x US-09-303-069-15 (1-2614)

Qy 6 IleGluAlaSerThrSerIleSerGlnLeuArg---GluSerTyrLeuAlaGluThr 24
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ATCCGCCGATCTACCTCGGGAGGAGCTGGTGGCTTCGCAGGATCCCTCGTGGCCACG 480

Qy 25 ArgHisArgGlyAsnAsnGluArgSerArgAlaGluPro----- 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CTGCAGCGCGCCCATCCCTCGGAGCGCCGCGAGCCGCCACTCTTCTCCCGGCTTCC 540

Qy 38 -----SerSerAsnProCysHisPheGlySer 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 ACACCCAAAGACCTCAGCGGCTGTAGCCGCGTGCACCCAGCCG-----CCGCCT 591

Qy 47 ProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeuProAspLeuSerAlaPhe 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 CCTAGTGTGGGGCAAA---TCTGGGACAGGCTGGGAGGCCGCCGGAAGCAGAGGGCCG 648

Qy 67 LeuSerGlnGluLeuAspGluSerValAsnLeuAlaArgLeuAlaIleAsnTyrAsp 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 GTGGCAGGACTGAACCGGGGAGGCGCCGAGGAGGATCAAGCGTCGGGACCAATTC 708

Qy 87 ProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeuSerProAspGlnMetLys 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 CCGCTAACCCAGGACGAGCCATCCAGGAGTGCAGG----- 744

Qy 107 HisSerProAsnLeuSerPheGluProAsnPheCysGlnAspAsnProArgSerProThr 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 745 ---AGCCCT-----GTGCGGCCCTACACCGCGGATCCCGGAGAGGAGGACA 789

Qy 127 SerSer-----LysSerProGlnGlnAlaLysArgProGlnTyrCys 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 790 AAAGCCCTCCGTCGCAAGCGGSAACCCCTGCTCAAGCGGTGCGC----- 837

Qy 142 SerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluLeu 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 -----TTTCTGCCCTGGGCCACTCCGGGAGTGGAGGAC--- 870

Qy 162 SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 871 -----TCTGTTCTGGCCCCCAACCTTGGAGAAAGAT 900

Qy 182 LysSerLysLeuGluSerGlnAsnLysValMetGln-----GluAsnSerSerSerPhe 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 901 AGCGGGGACCCGAGGCTGAGAAGAGGCTTCGCAGAGGACCTGAGGAGGATGGCCCTGG 960

Qy 200 SerAspLeuSerGluArgArgGluArgSerValProIleProAlaAspThr 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 961 GGGCCCTGGGACCGAGAGGACCCGCGAGC----- 990
```

```
QY 568 -----ProSerValGluGlnProProLysProLysLeuGluGlyValLeuValAsnHis 585
Db 1924 GCGGCGCGGACTAGCCATGCTCCCTTCCTCCCTAGCC-----CAT 1968
QY 586 AsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsnLeuProGluAsp 605
Db 1969 ACAGCACCC-----TGGGGTAACCCACCGGGC 1995
QY 606 AspLysGluSerGluAlaSer-SerGluAlaGlyValValThrArgGlnThrArgPr 625
Db 1996 CCTGTGGATCC-----TCCCTCCCAAGTGGATGGCTGTCACACAGGAGGCC 2049
QY 625 oAspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu---ProSerPh 644
Db 2050 CCAGAGG-ACTGAGTGTGGAGGATGGCCATGAGGGGTGCCAAGCTCCCTCGGTC 2108
QY 644 eProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
Db 2109 TCCCATAGGGAGCATCCAGCGAGTGCATGTGCTATGCTACAGGCCACT 2160
RESULT 7
US-09-134-250-15
; Sequence 15, Application US/09134250B
; Patent No. 6399753
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
; FILE REFERENCE: 05433/038001
; CURRENT APPLICATION NUMBER: US/09/134, 250B
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: US 08/795,868
; EARLIER FILING DATE: 1997-02-06
; EARLIER APPLICATION NUMBER: US 08/494,577
; EARLIER FILING DATE: 1995-06-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1806)
US-09-134-250-15
Alignment Scores:
Pred. No.: 1.6e-14 Length: 2614
Score: 326.00 Matches: 166
Percent Similarity: 35.10% Conservative: 79
Best Local Similarity: 23.78% Mismatches: 294
Query Match: 4.73% Indels: 160
DB: 4 Gaps: 25
US-09-818-990b-2 (1-1320) x US-09-134-250-15 (1-2614)
QY 6 IleGluAlaSerThrSerIleSerGlnLeuArg---GluSerTyrLeuAlaGluThr 24
Db 421 ATCCGCGGATCTACGTGCGGGAGGAGCTGGTGGCTGCGACGAGTCCCTGCTGCCACG 480
QY 25 ArgHisArgGlyAsnAsnGluArgSerArgAlaGluPro----- 37
Db 481 CTCAGCGCGGCCCATCCCTCCGGAGCGCGGAGCGGCCCATCTCTCTCCCGGCTTCC 540
QY 38 -----SerSerAsnProCysHisPheGlySer 46
Db 541 ACACCAAGACCTCAGCGGTGTGAGCGCGCGGTGCCACCCAGCCG-----CCGCGCT 591
QY 47 ProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeuProAspLeuSerAlaPhe 66
Db 592 CTAGTGGTGGCGGCAAA---TCTGGGAGGAGCGCTGGGAGGCCCGCGAGAGGGCGG 648
```

```
QY 67 LeuSerGlnGluLeuAspGluSerValAsnLeuAlaArgLeuAlaIleAsnTyrAsp 86
Db 649 GTGGGAGGACTGAACCGGGGAGAGCCCGCACGAGATCAAGCGTCGGGACCAATTC 708
QY 87 ProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeuSerProAspGlnMetLys 106
Db 709 CCCTTAACCAAGGAGAGACCATCCAGGAGTGCAG----- 744
QY 107 HisSerProAsnLeuSerPheGluProAsnPheCysGlnAspAsnProArgSerProThr 126
Db 745 ---AGCCCT-----GTCCCGCCCTACACCGCGGATCCCGCGGAGAGCAGACA 789
QY 127 SerSer-----LysGluSerProGlnAlaLysArgProGlnTyrCys 141
Db 790 AAGCCCTCCCGTCGTCGACGCGGAGAACCCCTGCTCAACGCGTGC----- 837
QY 142 SerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGluLeu 161
Db 838 -----TTTCTGCCCTGGGCCACTCCGGGAGTGGAGGAC--- 870
QY 162 SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis 181
Db 871 -----TCTGTTTCCCTCCCAACCTTGGGAGAGAAT 900
QY 182 LysSerLysLeuGluSerGlnAsnLysValMetGln-----GluAsnSerSerSerPhe 199
Db 901 AGAGCGGGAGCCCGAGCTGAGAGAGCTTCGACGAGGACCTGAGGAGGATGCCCTGG 960
QY 200 SerAspLeuSerGluArgGluArgGluArgSerValProIleProIleProAlaAspThr 219
Db 961 GGGCCCTGGGACCGCAGAGGAGCCCGCAGC----- 990
QY 220 ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGlu 239
Db 991 -----CAAGGCAAGAGTCCCGTGTCCG 1014
QY 240 GlnAlaAlaSerGluAlaGlyGlyAspThrProGlySerSerProSerSerLeu 259
Db 1015 CCTACTTCCCGGAGCTCGAGTCTCAGAGCTCTATGTGTCGCTGG----- 1065
QY 260 TyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGlu 279
Db 1066 -----GAGAGCGCCCTG---GAGGACCGCTGTTGAGATCCCTCTGCAAGATATGTTG 1116
QY 280 ValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGln 299
Db 1117 GTGGCGCGAGGAGCTGACGTGCTACTTAACTGATATCATCACCGCCACCCACCCCAA 1176
QY 300 ValArgTyrPyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGln 319
Db 1177 GTGTCCTGGAAAGAGTGGTCCATGTCACAGCGAGGCTGCTCTCTCATCCGGGCT 1236
QY 320 AlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyr 339
Db 1237 GAAGGTGAACGGCACACACTGCTGCTCAGAGAGGCCCGAGCTGCTGATGTTGGAGTAC 1296
QY 340 SerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGlu 359
Db 1297 ACAGCCACTGCCACCAACGAACTGGGCCAAGCTACCTGCTCTTCTTCCATGGTGAGAG 1356
QY 360 GlyValSerSerSerSerGluGlyAspProAsnLysGluGluMetAsnArgIleGln 379
Db 1357 CCTGGCGGCTCCATCCCTTTTCAGCAGGCCCATCATCTCTGATGAGGAGTACCTGAGC 1416
QY 380 LysProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValPro 399
Db 1417 CCCCCAGAGAGTTCACAGACCTGGGGAGACC----- 1449
QY 400 GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerPro 419
Db 1450 -----TGGCCCCGAAACCCCTACCATGAAGCTCAGTCCCGAGCCAG 1488
QY 420 ThrAsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLys 439
```

```
Db 1489 GATCATGATCTCCGACTCTTCTTCCAAG-----GCACCCCAACGTTTCAAGGTC 1539
Qy 440 MetLeuGlnAsnLeuSerAlaSerGluGlnLeuValPheGluCysArgValLys 459
Db 1540 TCACATGAGCAACCGTGGAGAGGTCAGATGTCATATGAGCATCCGTTGTCAG 1599
Qy 460 GlyAlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerPro 479
Db 1600 GGAGAGCCCAAGCTGTGGTTCTTGCTGAGGAATCGACAGCCCGTGGC-----CCA 1653
Qy 480 AspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGlu-----Ile 497
Db 1654 GAC-----CAGCGCGCTTTCAGAGGAGCGCGAGGGTGGGCTC 1692
Qy 498 CysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAla 517
Db 1693 TGGCGCTTGAGATCCCTGGCTGCTGAACGGGGCGATGCTGGTTCTACATGCAAGGCG 1752
Qy 518 SerAsnLysTyrGlyThrValSerIleAlaGlnLeuHisValArgGlyAsnGluAsp 537
Db 1753 GTCAACGAATATGGCTGCGCAGTCGAGCGCGCTGGAGGTCCGAGGGAG----- 1806
Qy 538 LeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGlu 557
Db 1807 ---TGAGCTCAGGGGGCCACCTGGCTGCTGCCCGCTACCTCCGAGCTGCACCCCTGCT 1863
Qy 558 ProGlnPro-----SerProHisSerGluPro----- 567
Db 1864 CAGCAGCTCTCGGACCTCGCTGTGTTTCATGCTGCTCTGCTGCCACAGACCCGCGCTC 1923
Qy 568 -----ProSerValGluGlnProProLysProLysLeuGluGlyValLeuValAsnHis 585
Db 1924 GCCGCGCGGACTAGCCCATGCTCCCTCCCTCCCTAGCC-----CAT 1968
Qy 586 AsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsnLeuProGluAsp 605
Db 1969 ACAGCACCC-----TGGGCTAACCCACCAGCGGC 1995
Qy 606 AspLysGlySerGluAlaSer-SerGluAlaGlyValValThrThrArgGlnThrArgPr 625
Db 1996 CCCTGTGATCC-----TCCCTCCCAAGTGGATGTGTGCTGCAGACCCAGGAGGCC 2049
Qy 625 oAspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu---ProSerPh 644
Db 2050 CCCAGAAGG-ACTGAGTGTGGGAAGGATGGCCATGAGGGGTCCCAAGCTCCCTCGTC 2108
Qy 644 eProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
Db 2109 TCCCATAGGAGGATCCAGCGAGTGTGCTATGCTGTATGCTGTACAGGCCACT 2160
```

RESULT 8

```
US-09-041-886-24
; Sequence 24, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Proapoptotic peptides, dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4342
; US-09-041-886-24
```

Alignment Scores:

Pred. No.:	3,03e-12	Length:	4608
Score:	300.00	Matches:	108
Percent Similarity:	41.13%	Conservative:	45
Best Local Similarity:	29.03%	Mismatches:	135
Query Match:	4.35%	Indels:	84
DB:	4	Gaps:	15

US-09-818-990b-2 (1-1320) x US-09-041-886-24 (1-4608)

```
Qy 959 GlySerProValThrPheThrCysLysIleValGlyLeuProValProLysValTyrTrp 978
Db 460 GGAGACACAGCTGTACTCAAGTGTGAAGTCTGGGAGGCCATGCCAACAATCCACTGG 519
Qy 979 PheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgGluGly 998
Db 520 CAGNAGAACCAACAGACCTGACTCCA-----ATCCCCAGGT 555
Qy 999 AspGly-----ThrCysSerLeuHisIleGluSerThrThrSerAsp 1012
Db 556 GACTCCCGAGTGTGTGCTTGCCTCTGTGGAGCATTCAGATCCAGCCGACTCCAACCGGG 615
Qy 1013 AspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGly 1032
Db 616 GACATTGGAATTTACCGATGCTCAGCTCGAAATCCAGCCAGC-----TCAAGAACAGGA 669
Qy 1033 HisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis 1052
Db 670 AAT-----GAAGCAGAAGTCAGATTTTATCAGATCCAGGACTGCAT 711
Qy 1053 ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArg 1072
Db 712 AGA-----CAG 717
Qy 1073 ProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeu 1092
Db 718 CTGTATTTCTGCAAGACCATCAATGTAGTACCCATTGAAGGAAAAGATGCTGCTCTG 777
Qy 1093 AspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnPro 1112
Db 778 GAATGTGTGTCTGGCTATCTCCACCAAGTTTACCTGG---TTACGAGCGGAGGAA 834
Qy 1113 ValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeu 1132
Db 835 GTCATCCAACCTCAGGCTCAAAAAGTATTTCTTTA-----TTGGTGTGAAGCAACTGCTT 888
Qy 1133 IleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThr 1152
Db 889 ATCTCCAATGTGACAGATGATGACAGTGAATGTATACCTGTGTGTGCACATATAAAAT 948
```

QY 1153 GlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaPro 1172
Db 949 GAGAAATATTAGTCCCTCGCAGAGCTCACAGTCTTGTT-----CCGCCA 993
QY 1173 ValLeuLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGlu 1192
Db 994 TGGTTTTTAATCATCTTCCCAACCTGTATGCTATGAAAGCATGATATTGAGTTTGAA 1053
QY 1193 CysArgValIleGlyMetProProValPheTyrTriPlyLysLysAspAsnGlu---Thr 1211
Db 1054 TGTACAGTCTCTGGAAGCCTGTGCCACTGTGAATGGATGATGAGATGTGGTC 1113
QY 1212 IleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeu 1231
Db 1114 ATTCCT-----AGTGATTATTTTCAGATAGTGGAGGAGCAACTTA 1155
QY 1232 LeuIleGlnProAlaLysLysSerAspAlaGlyTrpThrLeuSerAlaLysAsnGlu 1251
Db 1156 CGATACCTGGGTGGTGTGAAGTCCAGATGAGGCTTTTATCAATGTGTGGCTGAAATGAG 1215
QY 1252 AlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnIle 1271
Db 1216 GCTGGAATCCCGACAGCAGTGCACAGCTC-----ATTGTC 1251
QY 1272 Pro-----ProMetSerValArgProSer----- 1280
Db 1252 CCTAAGCCTGCAATCCCAAGCTCCAGTGTCTCCCTCGGCTCCAGAGATGTGGTCCCT 1311
QY 1281 -----GlySerArgTyrGlySerLeuThr-----SerLysGly 1291
Db 1312 GTCTTGTTTCCAGCGCATTTGTCCGTCTCAGCTGGCGCCACCTGCAGAGCGAAAGGG 1371
QY 1292 ---LeuAspIlePheSerAlaPheSerSerMetGlu 1302
Db 1372 AACATTCAAACTTTCAGGCTCTTTTCTCCAGAGAA 1407

RESULT 9

PCT-US94-05277-1
; Sequence 1, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zabrecky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 4608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 18q21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4342
; PCT-US94-05277-1

Alignment Scores: 3.03e-12 Length: 4608
Pred. No.: 300.00 Matches: 108
Score: 41.13% Conservative: 45
Percent Similarity: 29.03% Mismatches: 135
Best Local Similarity: 4.35% Indels: 84
Query Match: 5 Gaps: 15
DB:

US-09-818-990B-2 (1-1320) x PCT-US94-05277-1 (1-4608)
QY 959 GlySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrp 978
Db 460 GGAGACACAGTGTCTACTCAAGTGTGAAGTCATTTGGGAGCCCATGCCAACATCCACTGG 519
QY 979 PheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGly 998
Db 520 CAGAGAACCACACAGCTGACTCCA-----ATCCACAGT 555
QY 999 AspGly-----ThrCysSerLeuHisIleGluSerThrThrSerAsp 1012
Db 556 GACTCCCGAGTGGTGTCTTGCCTCTCGAGCATTCGAGATCAGCCGACTCCCAACCGGG 615
QY 1013 AspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGly 1032
Db 616 GACATTGGAAATTTACCGATGCTCAGCTCGAAATCCAGCCAGC-----TCAAGAACAGGA 669
QY 1033 HisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHis 1052
Db 670 AAT-----GAAGCAGAACTCAGAAATTTATCAGATCCAGGACTGCAT 711
QY 1053 ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArg 1072
Db 712 AGA-----CAG 717
QY 1073 ProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeu 1092
Db 718 CTGTATTTCTCAAGACACCATCAATGTAGTACCCATTCGAGGAAAGATGCTGCTCTG 777
QY 1093 AspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnPro 1112
Db 778 GAATGTGTGTCTGGCTATCTCCACCAAGTTTACCTGG---TTACAGGCGGAGGAA 834
QY 1113 ValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeu 1132
Db 835 GTCATCCAACTCAGGTCTAAAGAGTATTCCTTA-----TTGGGTGGAAGCAACTGCTCT 888
QY 1133 IleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThr 1152
Db 889 ATCTCAATGTGACAGATGATGACAGTGAATGTATACCTGTGTGTCACATATAAAAT 948
QY 1153 GlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaPro 1172
Db 949 GAGAAATATTAGTCCCTCGCAGAGCTCACAGTCTTGTT-----CCGCCA 993
QY 1173 ValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGlu 1192

```
Db 994 TGGTTTTAAATCATCTCTCCAACTGTATGCCATGAAGACATGGATATTGACGTTTGA 1053
Qy 1193 CysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGlu---Thr 1211
Db 1054 TGTACAGTCTCTGGAAAGCCTGTGCCACTGTGAATGGATGAAGATGGAGATGGTGC 1113
Qy 1212 IleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeu 1231
Db 1114 ATTCTCT-----AGTGATTATTTTCAGATAGTGGGAGGAACAACCTTA 1155
Qy 1232 LeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGlu 1251
Db 1156 CGGATACCTGGGGTGGTGAAGTCAAGTGAAGCTTTTATCAATGTGTGCTGAANAATGAG 1215
Qy 1252 AlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTTrpHisHisGlnIle 1271
Db 1216 GCTGGAATGCCAGACCAGTGCACAGCTC-----ATTGTC 1251
Qy 1272 Pro-----ProProMetSerValArgProSer----- 1280
Db 1252 CCTAAGCTGCAATCCCAAGCTCCAGTGTCTCCTCTTCGGCTCCAGAGATGTGTCCCT 1311
Qy 1281 -----GlySerArgTyrGlySerLeuThr-----SerLysGly 1291
Db 1312 GCTTTGGTTTCCAGCCGATTTTGTCCGTCTCAGCTGGCGCCACCCTGCAGAGAGCGAAAGGG 1371
Qy 1292 ---LeuAspIlePheSerAlaPheSerSerMetGlu 1302
Db 1372 AACATTCAAACTTTCACGGTCTTTTCTCCAGAGAA 1407

RESULT 10
US-09-484-970B-66
; Sequence 66, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 66
; LENGTH: 6814
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 199882.3
; NAME/KEY: unsure
; LOCATION: 1838, 5528
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-66

Alignment Scores:
Pred. No.: 2,06e-10 Length: 6814
Score: 278.50 Matches: 304
Percent Similarity: 31.25% Conservative: 172
Best Local Similarity: 19.96% Mismatches: 537
Query Match: 4.04% Indels: 514
DB: 4 Gaps: 69

US-09-818-990B-2 (1-1320) x US-09-484-970B-66 (1-6814)
Qy 47 ProSerGlyAlaAlaGluGlyGlyGlyGlyGlnAspAspLeuSer---Ala 65
Db 358 CCTAGTGAGCATTTGAA-----GACTTGGAAATTTAAATAATCTC 399
Qy 66 PheLeuSerGlnGluLeuAspGlu-----SerValAsn 77
Db 400 TATCTGTACAGAAATGAGATCCAGTCAATTGACAGGCAAGCATTTAAGGGACTTGCCTCT 459
```

```
Qy 78 LeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArg 97
Db 460 CTAGAGCAACTATACCTGCGACCTTTAATCAGATAGAACTTTGGAC----- 504
Qy 98 LysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPhe 117
Db 505 -----CCAGATTCTGTTCCAGCATCTCCGAAGCTC-----GAGAGCTATTT 546
Qy 118 CysGlnAspAsn-----ProArgSerProThrSerSerLysGluSerProGlnGluAla 135
Db 547 TTGCATTAACAACCGGATTACACATTTAGTTCCAGGGACATTTAATCACTTTGGAATCTATG 606
Qy 136 LysArgProGlnTyrCysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAla 155
Db 607 AAGAGATTGGCAGCTGGACTCAACACACACTTCACCTGCCACTGTGAATCTCTGTTGGCG 666
Qy 156 AspPheIleGluLeuSerSerLeuPheLysSerHisSerSer----- 170
Db 667 GATTGTGTAACCACTACGCGGAGTCGGGAACGCGCAGCAGCGGCCCATCTGTGTAATAT 726
Qy 171 ---LysArgIleArgProArgAla-----CysLys 179
Db 727 CCCAGACGCATCCAGGACGCTCAGTGGCAACCATCACCCCGGAAGAGCTGAATCTGTGA 786
Qy 180 AsnHisLysSerLysLeuGluSerGlnAsn---LysValMetGlnGluAsnSerSerSer 198
Db 787 AGCCCCGGATCACCTCCGAGCCCCCAGGACGACATGTGACCTCGGGGAACACCGTGTAC 846
Qy 199 PheSerAspLeuSerGluArgArgGlu-----ArgSerSerVal 211
Db 847 TTCACCTGCAGAGCGGAAGCAACCCCAAGCCTGAGATCATCTGGCTCGCAACAATAAT 906
Qy 212 ProIleProIleProAlaAspThrArgAspAsn----- 222
Db 907 GAGCTGAGCATGAAGACAGATTCCCGCTAAACTTGTGACGATGGGACCCCTGATGATC 966
Qy 223 -----GluValAsnHisAlaLeuGluGlnGlnGlnAlaLysArgArgGluAla 238
Db 967 CAGAACACACAGGAGACAGACAGGAGGTATCTACAGTGCATGGCAAGACAGTGGCGGA 1026
Qy 239 GluGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSer 258
Db 1027 GAGGTGAAGACGCAAGAGGTG-----ACC 1050
Qy 259 LeuTyrTyrGluGluProLeuGlyGlnPro-----ProArgPheThrGlnLysLeuArg 276
Db 1051 CTCAGGTAC-----TTCGGGTCTCCAGCTCGACCCCACTTTTGTAAATCCAGCCACAG 1101
Qy 277 SerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValValGlyIlePro 296
Db 1102 AATACAGAGGTGCTGGTTGGGAGAGCGTCAAGCTGAGTGCAGCCGCCACGAGCCACCCC 1161
Qy 297 ProProGlnValArgTrp---TyrCysGluGlyLysGluLeuGluAsnSerProAspIle 315
Db 1162 CCGCGCGGATCTCTCGGAGGAGGTGACCGCACACCTTGGCAGTTGACCGCGGGTG 1221
Qy 316 HistIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluAsp 335
Db 1222 AACATACGCTTCTGCGGGCTTTAC-----ATACAGACGTCGTACAGGGGGAC 1272
Qy 336 ThrGlyArgTyrSerCysPheAlaSerAsn-----IleTyrGlyThrAsp-Se 351
Db 1273 AGCGGAGGTATGCGTCTCTCGGACCAACAACATTTGACAGCGTCCATGCCCGCTTC 1332
Qy 351 rThrSerAlaGluIleTyrIleGluGlyValSerSerAspSerGluGly----- 368
Db 1333 ATCATCTCCAGGCTCTTCTCAGTTCACTGTGACGCTCAGGACAGAGCTGTTATTGAG 1392
Qy 369 -----AspProAsnLysGluLeuMetAsnArgIleGlnLys 380
Db 1393 GGCACAGCCGTGGATTTCACATGTGAAGCCAGGCGCAACCCGCGCCCTCATCGCCTGG 1452
```


Db	3303	GCTG-----GACGAGAACCTTCAGCGCCCATTCGACAAAGATCACTCCCCCT	3347
Qy	939	rgILyScysIleAlaProIlePheAspLysArgLeuLysHisPheArg---ValThrGI	958
Db	3348	TCACAAAGCTTTC-----TTCTCTCCCTTCGGATTGTGAATGA	3386
Qy	958	uGly-SerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrT	978
Db	3387	GGCGGCATCGATCCGCTTCTCAGGGGGCTGTTCGGG-----T	3425
Qy	978	rpPheLysaspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluG	998
Db	3426	GG-----CGGGAAATACGCTGCCCTCGCAGC-----TGTGAACACGGAGCTCAC	3473
Qy	998	lyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrT	1018
Db	3474	GGAGCGCTGTTCTCCATGGCACACACGGTGGCTC-----TGGA	3512
Qy	1018	hrIleMetAlaAlaAsnProGInGlyArgIleSerCysSerGlyHisLeuMetValGlnS	1038
Db	3513	CCTGGCGGCATCAACATCCAGCGGGCGGACACCGGATCCAC-----	3559
Qy	1038	erLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgV	1058
Db	3560	-----CCTACCAGACTACAGGGTCTACTGCAATCATATCGCGGCACACACGTTCTGAGGA	3614
Qy	1058	alGInGluArgaspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnA	1078
Db	3615	CCTGAAAATGAGATTAAAAACCTTGAGATCGGG-----AGAAACT	3656
Qy	1078	laProGlyAspMetValAlaHisGluGlyArgLeuCys-----ArgL	1092
Db	3657	GAAAAGTGTATGGCTCGACACTCAACATCGACCTGTTCCGGCGCTCGTGGTGAGGA	3716
Qy	1092	euAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyClnP	1112
Db	3717	CCTGGTGCCTGGCAGCGCTGGCGCCCACTGATGTCTTC-----	3760
Qy	1112	roValLeuProAspAlaSerHisLysMetLeuValArgGlu-----ThrGlyV	1128
Db	3761	-----TCAGCACACAGTTCAAGCGCTCGCAGATGGGACAGAGTTGTGGTA	3806
Qy	1128	alHisSerLeuLeuIleAspPro-----LeuThrGlnArgAspA	1141
Db	3807	TGAGAACCTCTGGGTGTTCTCTCCCGCCGACGTGACTCAGATCAAGCAGACGCTGCTGGC	3866
Qy	1141	laGlyThrTrpLysCysIleAlaThrAsnLysThrGly-----GlnAsnSerP	1157
Db	3867	CAGGATCCTATCGCACACGCGGACACATCACC GGGTGCGAGCAGCAGCTGTTCAGGT	3926
Qy	1157	heSerLeuGluLeuSerValAlaLysGluValLysLysAlaProValIleLeuGluL	1177
Db	3927	GGCGGAGTCCCTCAGCGGTACGGCA---GCTGTGACGAGATCCCCA-----GGGT	3974
Qy	1177	ysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleG	1197
Db	3975	GGACCTCCGGGTGTGC-----AGGACTGCTGTGAAGACTGTAG	4013
Qy	1197	ly-----MetProProProValPheTyrTrpLysLysAspAsnGluThrI	1212
Db	4014	GACACGGGGCAGTTCAATATGCTTTTTCCTATC-----	4045
Qy	1212	leProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuL	1232
Db	4046	-----ATTTCGAGGACGAGCGTCTCTTGAGTTCTACGTACCAGGA	4085
Qy	1232	euIleGlnProAlaLysLysSerAspAlaGlyTyrTrpLysLysAspAsnGluL	1252
Db	4086	GGACAAGCCGACCAAGAAAACAGACCGGAAAATACCCAGTG-----TTGGGAGACA	4139
Qy	1252	laGlylleValSerCysThrAla---ArgLeuAspIleTyrAlaGlnTrpHisHisGlnI	1271
Db	4140	GGGGGAACATCTCAGCAACAGACACCTCAGCTTCGACGACGCTCAGATCATCAT	4192

Qy	1271	leProProMetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysG	1291
Db	4193	-----CTGGGACAAATGACTTCAGAGA	4214
Qy	1291	lyLeu	1292
Db	4215	GTTCG	4219
RESULT 11			
US-07-757-022B-51			
; Sequence 51, Application US/07757022B			
; Patent No. 6433142			
; GENERAL INFORMATION:			
; APPLICANT: Gesner, Thomas G.			
; APPLICANT: Clark, Stephen C.			
; APPLICANT: Turner, Katherine			
; APPLICANT: Hewick, Rodney M.			
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors			
; NUMBER OF SEQUENCES: 143			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Genetics Institute, Inc.			
; STREET: 87 CambridgePark Drive			
; CITY: Cambridge			
; STATE: Massachusetts			
; COUNTRY: U.S.A.			
; ZIP: 02140			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/07/757,022B			
; FILING DATE: 19910910			
; CLASSIFICATION: 530			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 07/643,502			
; FILING DATE: 18-JAN-1991			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 07/546,114			
; FILING DATE: 29-JUN-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 07/457,196			
; FILING DATE: 29-DEC-1989			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 07/390,901			
; FILING DATE: 08-AUG-1989			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Cserr, Luann			
; REGISTRATION NUMBER: 31,822			
; REFERENCE/DOCKET NUMBER: GI 5190			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (617)876-1170			
; TELEFAX: (617)876-5851			
; INFORMATION FOR SEQ ID NO: 51:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 4092 base pairs			
; TYPE: NUCLEIC ACID			
; STRANDEDNESS: double			
; TOPOLOGY: unknown			
; MOLECULE TYPE: cDNA			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 1..4089			
US-07-757-022B-51			
Alignment Scores:			
Pred. No.: 1.04e-10 Length: 4092			
Score: 278.00 Matches: 239			
Percent Similarity: 31.41% Conservative: 115			
Best Local Similarity: 21.21% Mismatches: 397			
Query Match: 4.03% Indels: 379			


```
Db 1469 CCACCACCTACCAAGGAGCGCTGCACCCACCACTACCAAGTCTGCACCCACCACTCCCAAGG 1528
Qy 753 hrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerValGlnThrL 773
Db 1529 AGCCTTCACCCACCAACCAAG-----1551
Qy 773 ysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProProGlyProThrGluProT 793
Db 1552 -----GAGCCTGCACCCACCACTCCCAAGGAGCGCTG 1582
Qy 793 hr-ProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysVal 812
Db 1583 CACCACCAACCC-----CCTACCCCAAGGAGCGCTG 1630
Qy 813 SerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSerSer 832
Db 1631 CACCCA---CCTC---CCAGGACCTGCACCCACCAACCAAGGAGCTGCACCCA 1684
Qy 833 ValLeuPro---SerLeuProAlaIleProProThrAsnAla---Met***LeuProArg 850
Db 1685 CGCTCCCAAGAGCGCTGCCCACTACCCCAAGGAGACTGCACCCACCAACCCCAAGA 1744
Qy 851 SerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnPro 870
Db 1745 AGCTCAGCCCAACCAACCCCGAGAAGCTCGCACCCACCAACCCCTGAGAAGCGCGCACCA 1804
Qy 871 -----ValAsnAspAsnIleArgGluThrLysAsnAlaValIleArg 885
Db 1805 CCACCCCTGAGAGCTGCACCCACCCCTGAGAGCGCCACACCCACCAACCCCTGAGG 1864
Qy 886 AspLeuGlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGluTyrLysIle 905
Db 1865 AGCCTGCTCCCACTCCCA-----1885
Qy 906 SerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProVal 925
Db 1886 -----AGGCGCGCTGCCAACACCCCTAAGGAGCGCTGCCAACT-ACCCCTAAG 1935
Qy 926 AspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCysIleAlaPro 945
Db 1936 GAGCTGCTCCCACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGACTGCTCA 1995
Qy 946 IlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
Db 1996 ACTACCCCTAAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCCACT-----2049
Qy 966 CysLysIleValGlyIleProValProLysValTyrTyrPheLysAspGlyLysGlnIle 985
Db 2050 -----CCCAAGAAGCTGCCCCCAAG-----2070
Qy 986 SerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHis 1005
Db 2071 -----GAGCTTGACCCCAACCAACCAAGGAG 2097
Qy 1006 IleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMetalAlaAsnProGln 1025
Db 2098 CCCATCCACCACTCTGCACAGCCGCTCCAACTAC-----CCTAAG 2142
Qy 1026 GlyArg-IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArgLe 1045
Db 2143 GGGAGCTGCTCCACTACCCCTAAGGAGCGCTGCTCCAA---CTACCCCTAAGGAGCGCTGCT 2199
Qy 1045 uThrSerAlaGlnGlnSerHisArgGlyArgSerArgValGlnGluArgAspLysGluPr 1065
Db 2200 CCAA-----CTACCCCTAAGGAGCTGCTCCAACTACCTCAAGGAACCTGCACCC 2250
Qy 1065 oLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHi 1085
Db 2251 ACTACTCCCAAGAGCGCTGCCCAAGGAGCTTGACCCACCAACCAAGGGGCCCA 2310
Qy 1085 sGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluLeuTh 1105
Db 2311 T-----CCACCACCTCTGAC 2325
```

```
Qy 1105 rTrpLeuLeu-----AsnGlyGlnPr 1112
Db 2326 AAGCTGTCTCAACTACACTAAGGAGACTGCTCCAACTACCCCAAGGAGCTGCACCC 2385
Qy 1112 oValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLe 1132
Db 2386 ACTACCCCAAGAAGCGCTGCTCCAACTACTCTGAGACAC-----CT 2427
Qy 1132 uIleAspProLeu 1136
Db 2428 CCTCAACCACTT 2440
```

```
RESULT 12
US-07-757-022B-103
; Sequence 103, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990.
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserfi, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3420 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3420
US-07-757-022B-103
```

```
Alignment Scores: 1.32e-10 Length: 3420
Pred. No.:
```

Score: 275.00 Matches: 229
 Percent Similarity: 31.03% Conservative: 103
 Best Local Similarity: 21.40% Mismatches: 376
 Query Match: 3.99% Indels: 365
 DB: 4 Gaps: 43

US-09-818-990B-2 (1-1320) x US-07-757-022B-103 (1-3420)

QY 116 AsnPheCysGlnAsp-----AsnProArgSerProThrSerLysGluSerProGln 133
 Db 304 AGTTCTGTGCAGAAAGTGCATATCCACATCACCACCATCTTCAAAGAAGACCTCCA 363
 QY 134 GluAlaLysArgProGlnTyrCysSerGluThrGlnSerLysLysValPheLeuAsnLys 153
 Db 364 -----CCTTCAGGAGCATCTCAAACC----- 384
 QY 154 AlaAlaAspPheIleGluGluLeuSerSerLeuPheLysSerHisSerLysArgIle 173
 Db 385 -----ATCAAAATCAACACCAACCGTTACCC 411
 QY 174 ArgProArgAlaCysLysAsnHisLysSerLysLeuGluSerGlnAsn----- 189
 Db 412 AAACCCCAACAAAGAGACAGTAAGAGTTATAGAAATCAGAGAAATAACAGAGAA 471
 QY 190 -----LysValMetGlnGluAsnSerSerPheSerAsp----- 201
 Db 472 CATTTCTGTTTGAATAATCAAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 531
 QY 202 -----LeuSerGluArgGluArgGluArgSerSerValProIleProIlePro 216
 Db 532 ACATTTTGGAAAAATCAAGTCTTCCAAAATTCAGCTGCTAATAGAGAAATACAGAGAAA 591
 QY 217 AlaAspThrArgAspAsnGluValAsnHisAlaLeuGluGlnGluAlaLysArgArg 236
 Db 592 CTCAAAGTAAAGATACAGAGACAGACAGTAAGAAAGAACTACCCCAACACCA 651
 QY 237 GluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAsp-----ThrThrProGly 253
 Db 652 GTTGTAGATGAAGCTGGAAGTGGATGGACAATGGTGACITCAAGTCAACACTCTCTGAC 711
 QY 254 SerSerProSerSerLeuTyrTyrGluGluProLeuGlyGlnProProArgPheThrGln 273
 Db 712 ACGTCTACCAACCCAA-----CACAAATAAGTACAGCATCTCCCAAGATCACACA 762
 QY 274 LysLeuArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIleValVal 293
 Db 763 GCAAAACCAATAATCCAGACC----- 786
 QY 294 GlyIleProProGlnValArgTyrTyrCysGluGlyLysGluLeuGluAsnSerPro 313
 Db 787 AGTCTTCCACCT----- 798
 QY 314 AspileHisIleValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGlu 333
 Db 799 -----AAT 801
 QY 334 GluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSer 353
 Db 802 TCTGATACATCTAAGAGACCTCTTTGACAGTGAATAAGAGACACAGATGTAACCTAAA 861
 QY 354 AlaGluIleTyrIleGluGlyValSerSerSerAsp-----SerGluGlyAspProAsnLys 372
 Db 862 GAAACTACTACAACAAATAACACACTTCAACTGATGGAAAAGAGAGACTACTTCCGCT 921
 QY 373 GluGluMetAsnArgIleGlnLysProAsnGluValSerSerProProThrThrSerAla 392
 Db 922 AAAGACACAAAGATATAGAGAAAACATCTGCTAAAGATTTAGCACCCACATCTAAAGTG 981
 QY 393 ValIleProProAlaValProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThr 412
 Db 982 CTGCTAAACCT-----ACACCCCAAGCTGAACACTACAACCAAGCCCTGCTCTCACCACT 1038
 QY 413 IleGlnGlnCysGlnSerProThrAsnTyrLeuGlnGlnGlyLeuAspGlyLysProIleIle 432

Db 1039 CCCAAGGAG---CCACGCCCCACCATCTCCCAAGGAGCTGCTCTACCCACCCAAAGAG 1095
 QY 433 AlaAlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuVal 452
 Db 1096 CCCACACCTACC----- 1107
 QY 453 ValPheGluCysArgValLysGlyAlaProSer-ProLysValGluTrpTyrArgGluGlu 472
 Db 1108 -----ACCATCAAGTCTGCACCCACCAACC----- 1132
 QY 472 yThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAl 492
 Db 1133 -----CCAAGGAGCCTGCA 1146
 QY 492 aGluPro-GluGluIleCysThrLeuValIleAlaGluValPheAlaGluAspSerGlyC 512
 Db 1147 CCCACCCACCAAGTCTGCACC----- 1169
 QY 512 ysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerIleAlaGlnLeuHisV 532
 Db 1170 -----CACCACT 1176
 QY 532 alArgGlyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrA 552
 Db 1177 CCCAAGGAGCCTGCACCCACCAACCAAGGAGC-CTGCAC----- 1216
 QY 552 snLeuAlaIleGluProGlnProSerProProHisSerGluProProSerValGluG 572
 Db 1217 -----CCACCACTCCCAAGGAGCCTGCACCCACCAACCAAGGAGCCTGC 1262
 QY 572 lnProLysProLysLeuGluGlyValLeuValAsnHisAsnGluProProArgSerS 592
 Db 1263 ACCCACCAACCAAGTCTGCACCCACCACTC-----CCAAGGAGC----- 1303
 QY 592 erArgIleGlyLeuArgValHisPheAsnLeuProGluAspAspLysGlySerGluAla 612
 Db 1304 -----CTGCACCC-ACCACCCCAAG---AAGCTGCCCCCAACTACCC 1342
 QY 612 erSerGluAlaGlyValValThrThrArgGlnThrArgProAspSer***GlnGluArgp 632
 Db 1343 CCAAGGAGCCTGCACCCACCACTCCCAAGGAGCCTACACCCACCACTCCCAAGGAG---- 1398
 QY 632 heAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProProV 652
 Db 1399 --CCTGCACCCACCAAGGAGCCTGCACCCACCACTCCC---AAAGAGCCTGCACCCA 1453
 QY 652 alLeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeu 672
 Db 1454 CTGCCCCCAAGAGCCTGCCCAACTACCCCC----- 1485
 QY 672 euGluGlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe*** 692
 Db 1486 -----AAGGAGCCTGCACCCACCACTCCCAAGGAG----- 1515
 QY 692 etThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysA 712
 Db 1516 -----CTGCACCCACCACTCCCAAGGAG-----AAGG 1537
 QY 712 laProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnT 732
 Db 1538 AGCCTTCCACCCACCACT-----CCCAAGGAGCCTGCACCCACCACTCCCAAGGAG 1582
 QY 732 hrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSer-----SerT 750
 Db 1583 AGTGTGACCCCACTACCAAGGAGCCTGCACCCACCACTACCAAGTCTGCACCCACCA 1642
 QY 750 hrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerV 770
 Db 1643 CTCCAAGGAGCCTTCAACCCACCACTCCCAAG----- 1674
 QY 770 alGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProGlyProT 790

Db 1745 AGGAGCCTGCACCCA---CCACTC---CAAAGGAACCTGCACCCACCACCCACCAAGAAGC 1798
Qy 830 LeuSerSerValLeuPro---SerLeuProAlaIleProProThrAsnAla---Met*** 847
Db 1799 CTGCACCCACCGCTCCCAAGAGCCTGCCCACTACCACTACCCCAAGAGGACTGCACCCACCA 1858
Qy 848 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer 867
Db 1859 CCCCACCAAGAGCTCAGCCACCCACCCCGAGAGCTGCACCCACCCACCCCTGAGAGC 1918
Qy 868 ProGlnPro-----ValAsnAspAspAsnIleArgGluThrLysAsnAla 882
Db 1919 CCGCACCACCCACCCCTGAGGAGCTGCACCCACCCACCCCTGAGGAGCCACACCCACCA 1978
Qy 883 ValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGlu 902
Db 1979 CCCCAGGAGCCTGCTCCACCACTCCCA----- 2008
Qy 903 TyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArg 922
Db 2009 -----AGCAGCGGCTCCCAACACCCCTAAGGAGCCTGCTCCCACT- 2049
Qy 923 ThrProValaspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCys 942
Db 2050 ACCCCTAAGGAGCCTGCTCCCACTACCCCTAAGGAGCCTGCTCCCACTACCCCTAAGGAG 2109
Qy 943 IleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProVal 962
Db 2110 ACTGCTCCCACTACCCCTAAGGAGCCTGCTCCCACTACCCCTAAGGAGCCTGCTCCCACT 2169
Qy 963 ThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysaspGly 982
Db 2170 ACT-----CCCAAGAGCCTGCCCCACAG----- 2193
Qy 983 LysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCys 1002
Db 2194 -----GAGCTTGACCCACCCACC 2211
Qy 1003 SerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAla 1022
Db 2212 ACCAAGGAGCCCATCCCACTCCCACTCTGACAAAGCCCGCTCCCACTACC----- 2259
Qy 1023 AsnProGlnGlyArg-IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleArg 1042
Db 2260 ---CCTAAGGGAGCTGCTCCCACTACCCCTAAGGAGCCTGCTCCCA---CTACCCCTAAG 2313
Qy 1042 gSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGluArgAs 1062
Db 2314 GAGCCTGCTCCAA-----CTACCCCTAAGGGAGCTGCTCCCACTACCCCTCAAGGAA 2364
Qy 1062 pLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMe 1082
Db 2365 CTGCACCCACTACTCCCAAGAGCCTGCCCAAGGAGCTTGACCCACCCACCCACCCAAAG 2424
Qy 1082 tValAlaHisGluGlyArgLysArgLysValSerGlyLeuProProPr 1102
Db 2425 GGGCCACAT-----CCACC 2439
Qy 1102 oGluLeuThrTrpLeuLeu-----As 1109
Db 2440 ACCTCTGACAAGCCTGCTCCCACTACACCTAAGGAGACTGCTCCCACTACCCCAAGGAG 2499
Qy 1109 nGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHi 1129
Db 2500 CTGCACCCACTACCCCAAGAGCCTGCTCCCACTACTCTCTGAGACAC----- 2548
Qy 1129 sSerLeuLeuAspProLeu 1136
Db 2549 -----CTCCTCCCAACCACTT 2563

RESULT 14

US-07-757-022B-1

; Sequence 1, Application US/07757022B

; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/757,022B
; APPLICATION NUMBER: 19910910
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5008 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4212
; US-07-757-022B-1

Alignment Scores:
Pred No.: 2,33e-10 Length: 5008
Score: 275.00 Matches: 229
Percent Similarity: 31.03% Conservative: 103
Best Local Similarity: 21.40% Mismatches: 376
Query Match: 3.99% Indels: 365
DB: 4 Gaps: 43

US-09-818-990B-2 (1-1320) x US-07-757-022B-1 (1-5008)
Qy 116 AsnPheCysGlnAsp-----AsnProArgSerProThrSerLysGluSerProGln 133
Db 304 AGTTTCTGTGCAGAGTGTCATAATCCACATCCACCATCTTCAAGAAAGACCTCCA 363
Qy 134 GluAlaLysArgProGlnTyrCysSerGluThrGlnSerLysValPheLeuAsnLys 153
Db 364 -----CCTTCAGGAGCATCTCAAC----- 384

Db	1147	 CCCACCACCACCAAGTCTGCAC-	1169
Qy	512	ysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisV	532
Db	1170	-----CACCAC	1176
Qy	532	alArgGlyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrA	552
Db	1177	CCCAAGGAGCCTGCACCCACCACCACCAAGGAGC-CTGCAC	1216
Qy	552	snLeuAlaAlaIleGluProGlnProSerProProHisSerGluProProSerValGluG	572
Db	1217	-----CCACCACCTCCCAAGGAGCCTGCACCCACCACCAAGGAGCCTGC	1262
Qy	572	lnProProLysProLysLeuGluGlyValLeuValAsnHisAsnGluProArgArgSerS	592
Db	1263	ACCACACCACCAAGTCTGCACCCACCAC	1303
Qy	592	erArgIleGlyLeuArgValHisPheAsnLeuProGluAspLysGlySerGluAlaAs	612
Db	1304	-----CTGCACCC-ACCACCCCAAG--AAGCCTGCCCACTACCC	1342
Qy	612	erSerGluAlaGlyValValThrThrArgGlnThrArgProAspSer**GlnGluArgP	632
Db	1343	CCAAGGAGCCTGCACCCACCACCTCCCAAGGAGCCTACACCCACCACCTCCAAGGAG	1398
Qy	632	heAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProProPro	652
Db	1399	--CTGCACCCACCACCAAGGAGCCTGCACCCACCACCTCCC--AAAGAGCCTGCACCA	1453
Qy	652	alLeuAlaLysProLysLeuAspSerThrGlnLeuGlnLeuHisAsnGlnValLeuL	672
Db	1454	CTGCCCCCAAGCCTGCCCACTACCCC	1485
Qy	672	euGluGlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe**M	692
Db	1486	-----AAGGAGCCTGCACCCACCACCTCCAAGGAG	1515
Qy	692	etThrValLeuAsnSerAsnAlaProProAlaValThrThrSer**LysGlnValLysA	712
Db	1516	-----CCTGCACCCACCACCC	1537
Qy	712	laProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnT	732
Db	1538	AGCCTTACCACCACT------CCCAAGGAGCCTGCACCCACCACCA	1582
Qy	732	hrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSer-----SerT	750
Db	1583	AGTCTGCACCCACCACCTACCAAGGAGCCTGCACCCACCACCTACCAAGTCTGCACCA	1642
Qy	750	hrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerV	770
Db	1643	CTCCCAAGGAGCCTTACCACCAACCAACCAAG	1674
Qy	770	alGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProGlyProT	790
Db	1675	-----GAGCCTGCACCCACCACCTCCCA	1696
Qy	790	hrGluProThr-ProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnPro	809
Db	1697	AGGAGCCTGCACCCACCACCC-CCAAGAGGCTGCC-CAACTACCCCA	1744
Qy	810	ArgCysValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPhe	829
Db	1745	AGGAGCCTGCACCA-CCACTC-CCAAGAGCCTGCACCCACCACCAAGAGC	1798
Qy	830	LeuSerSerValLeuPro---SerLeuProAlaIleProProThrAsnAla---Met***	847
Db	1799	CTGCACCCACCGCTCCCAAGAGCCTGCCCACTACCCCAAGGAGACTGCACCCACCA	1858
Qy	848	LeuProArgSerValProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer	867

```
Db 1859 CCCCCAAGAAGCTCAGCCCCACCCACCCCGGAGAGCTCGCCACCCACCACCCCTGAGAAAGC 1918
Qy 868 ProGlnPro-----ValAsnAspAspAsnIleArgGluThrLysAsnAla 882
    ||| |||
Db 1919 CCGCACCACACCCCTGAGAGCTCGCAGCCACCACCCCTGAGGAGCCACACCCACCA 1978
Qy 883 ValIleArgAspLeuGlyLysIleThrPheSerAspValArgProAsnGlnGlnGlu 902
    :||| |||
Db 1979 CCGCTGAGGAGCTGCTCCCACTCCCA----- 2008
Qy 903 TyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArg 922
    :||| |||
Db 2009 -----AGCAGCGGCTCCCAACACCCCTAAGGAGCGCTCTCCAAC- 2049
Qy 923 ThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCys 942
    :||| |||
Db 2050 ACCCTTAGGAGCGCTGCTCCCACTACCCCTAAGGAGCGCTGCTCCAACCTTAAGGAG 2109
Qy 943 IleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProVal 962
    :||| |||
Db 2110 ACTGCTCCAACCTACCCCTAAAGGAGCTGCTCCAACCTACCCCTCAAGGAACCTGCACCCACT 2169
Qy 963 ThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysAspGly 982
    :||| |||
Db 2170 ACT-----CCCAAGAAGCTGCCCCCAAG----- 2193
Qy 983 LysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCys 1002
Db 2194 -----GAGCTTGACCCACCACCTACC----- 2211
Qy 1003 SerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAla 1022
    :||| |||
Db 2212 ACCAAGGAGCCACATCCACCACCTCTGACAAGCCCGCTCCAACCTACC----- 2259
Qy 1023 AsnProGlnGlyArg-IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleAr 1042
    :||| |||
Db 2260 ---CCTAAGGGAGCTGCTCCAACCTACCCCTAAGGAGCGCTGCTCCA-----CTACCCCTAAG 2313
Qy 1042 gSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGluArgAs 1062
    :||| |||
Db 2314 GAGCTGTCTCAA-----CTACCCCTAAGGGAGCTGCTCCAACCTACCCTCAAGGAA 2364
Qy 1062 pLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMe 1082
    :||| |||
Db 2365 CCGTGCACCCACTACTCCCAAGAAGCGCTGCCCCCAAGGAGCTTGCAACCCACCACCAACCAAG 2424
Qy 1082 tValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProPr 1102
    :||| |||
Db 2425 GGGCCCACT-----CCACC 2439
Qy 1102 oGluLeuThrTrpLeuLeu-----As 1109
    :||| |||
Db 2440 ACCTGTGACAAGCTGCTCCAACCTACACCTAAGGAGACTGCTCCAACCTACCCCAAGGAG 2499
Qy 1109 nGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHI 1129
    :||| |||
Db 2500 CTGCACCCACTACCCCAAGAAGCGCTGCTCCAACCTACTCTCTGAGACAC----- 2548
Qy 1129 sSerLeuLeuIleAspProLeu 1136
    :||| |||
Db 2549 -----CTCTCTCAACCACTT 2563

RESULT 15
US-09-877-730-31
; Sequence 31, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
```

```
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 3874
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-877-730-31

Alignment Scores:
Pred. No.: 2,63e-10 Length: 3874
Score: 272.00 Matches: 241
Percent Similarity: 29.06% Conservative: 153
Best Local Similarity: 17.77% Mismatches: 433
Query Match: 3.95% Indels: 529
DB: 4 Gaps: 47

US-09-818-990B-2 (1-1320) x US-09-877-730-31 (1-3874)
Qy 229 GluGlnGlnGluAlaLysArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGly 248
    ||| ||| :||| ||| :||| ||| :||| |||
Db 23 GAGCACCAGCCTCGCCCGCAGAGACGCTCGCCACCGCGGTGCCGAGCGGGAGGG 82
Qy 249 AspThrThr-----ProGlySerSerProSerSer 258
    :||| ||| :||| |||
Db 83 AGCCAGGCTGCTGCGACCGCCAGAGCAGAGAGCGCCGCGGGCGGGCGGAGCG 142
Qy 259 LeuTyrTyrGluGluProLeuGlyGlnProArgPheThrGlnLysLeuArg----- 276
    :||| ||| :||| |||
Db 143 -----GACAGGGGCTCTGGCGCGCGGGAGCATGCCCGCGGCTACGCTGAATG 193
Qy 277 -----SerArgGluValProGluGly----- 283
    :||| ||| :||| |||
Db 194 GCGCTCTCTGCGACCCCTCGCCCGGCTCGGACCGCGGGGATGCTCTCGCGCGCTC 253
Qy 283 ----- 283
Db 254 CTGCTCTGCTGMRGCTCAGTCTTGTCCAGAGTGTGCTTTAGCGAACTGCTCTTTT 313
Qy 284 -----ThrArg-----ValGlnLeuAspCysIle 291
    :||| ||| :||| |||
Db 314 GTAAAGAACACAGGATGTAACTGTCAAGAAGAGCCAGCTGTTTAGATTGCCAG 373
Qy 292 ValValGlyIleProProGlnValArgTrpTyrCysGluGlyLysGluLeuGluAsn 311
    :||| ||| :||| |||
Db 374 GCTCAGGAGAGTTCCTATTAAAGTTCATGTTGAAAATGGAGCAAAATGCTGAA 433
Qy 312 SerProAspIleHisIleValGlnAlaGlyAsnLeuHisSerLeuThrIle-----Ala 329
    :||| ||| :||| |||
Db 434 AATAACGGATCGAGGTCTTTTCTTAACGGCTCTTTATATACATCAGTGAGTGAAGGCGAG 493
Qy 330 GluAlaPheGluGluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThr 349
    :||| ||| :||| |||
Db 494 CGAGGAGAGCAGCTCCGATGAAGGATTTTATCAGTGTGCGAATGAACAAATATGGAGCC 553
Qy 350 Asp-----SerThrSerAlaGluIleTyrIleGluGly 360
    :||| ||| :||| |||
Db 554 ATTCTTAGTCAAAAAGCTCATCTTGCCTTATCAACTATTCTGCAATTGGAAGTCCAGCCA 613
Qy 361 ValSerSerSerAspSerGluGlyAspProAsnLys----- 372
    :||| ||| :||| |||
Db 614 ATTTCCACTGAGGTCCAGGAAGTGGAGTGTCTCGATTGTCATCAAGATTTTCATCCAC 673
Qy 373 -----GluGluMetAsnArgIleGlnLysPro-----AsnGlu 383
    :||| ||| :||| |||
Db 674 CCTCTCGCATATACATGGAGTTCAATCGGACAACTCTTACCTATGACTATGAGCAGG 733
```


Db 2345 GTCCCTCTCTCCACACACACCC----- 2365
 Qy 989 AsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGlu--- 1007
 Db 2366 --CACCATCTCTATGGAGGCTAACACCTCATCTTCCTGCTGCTGAGGAGG 2422
 Qy 1008 ---SerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGly 1026
 Db 2423 CCTGCATTACCGCTGCACAAATCATTAACATACACCATC---CGCTGTAACTCTGTTGCG 2479
 Qy 1027 ArgIleSerCysSer-----GlyHisLeuMetValGln 1037
 Db 2480 CTGCAGATGCTTCTTGGTGTCTACCTTCAACATCAAGAACTCACATGTGTTTCAA 2539
 Qy 1038 SerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArg 1057
 Db 2540 GGTCTAGACCAACCAATAGCAATTTGCCGTTCCGATTACAT----- 2584
 Qy 1058 ValGlnGluArgAspLysGluProLeuGlnGluArgPheArgProHisPheLeuGln 1077
 Db 2585 ---GTGGATCAGCTTCTCCAGTCTCTGGAGCCCTGTAGTCTACCATCTACTCTCCAGAA 2641
 Qy 1078 AlaPro---GlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysVal 1096
 Db 2642 GCACACAGCCGCCACAGTTGGAGTAAAGTGACATTAATAGAGGATGACACATGCCCTG 2701
 Qy 1097 SerGlyLeuProProGlu-----LeuThrTrpLeuLeu 1108
 Db 2702 GTTCTTGGAAACCCCTGATGGCCAGAAACAGTTGTGACCGCTATCTATTATAT 2761
 Qy 1109 AsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyVal 1128
 Db 2762 GCATCTAGGAAGGCTCGATTGCGAGAGAGTGCAGGTCTTACACCGTGAAGGGCAATA 2821
 Qy 1129 HisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAla 1148
 Db 2822 ACCATGCTTTGTTAGAAAACTTTGGTAGCAGGAAATGTGTACATTGTCAAGATATCTGCA 2881
 Qy 1149 ThrAsnLysThrGlyGlnAsnSerPhe-----SerLeuGluLeuSerValValAlaLys 1166
 Db 2882 TCCAAATGAGTGGGAGAGACCCCTTTTCAATCTTGTGAGCTGGCAGTACTTCCAAAG 2941
 Qy 1167 GluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCys----- 1181
 Db 2942 GAAACCTCTGAATCAATCAGAGGCCCAAGCGTTTAGATTCTGCTGATGCCAAAGTTTAT 3001
 Qy 1182 -----GlyValProGluGlyHisProVal 1189
 Db 3002 TCAGGATATTACCATCTGGACCAAAAAATCAATGACTGGCATTGCTGTAGTGTGGCATA 3061
 Qy 1190 ArgLeuGluCysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsn 1209
 Db 3062 GCCTTGACCTGCATCTCTCATC----- 3082
 Qy 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAla 1229
 Db 3082 ----- 3082
 Qy 1230 CysLeuLeuIle-----GlnProAlaLysSerAspAlaGlyTrpTyrThr 1245
 Db 3083 TGTGTTCTCATCTGTATATACCGAAGTAAAGCCAGGAAATCATCTGCT-----TCC 3133
 Qy 1246 LeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAla 1265
 Db 3134 AAGCGGCACAGAAATGGAAC----- 3154
 Qy 1266 GlnTrpHisGlnIleProProMetSerValArgProSerGlySerArgTyrGly 1285
 Db 3155 -----CAACAGTTACCTCTGCTACCGAGCTCTTAGCTAGTGGAAATGAGGTAGGA 3205
 Qy 1286 SerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerSerMetGluSerThrMet 1305
 Db 3206 -----AAGAACTTGGAA-----GGAGCTGTAGGAAATGAAGAAATCTTTAATG 3247

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 30, 2002, 20:59:00 : Search time 93 Seconds
(without alignments)
5466.155 Million cell updates/sec

Title: US-09-818-990B-2

Perfect score: 6890

Sequence:

1 MDDSTEAISTSISQLRESY.....MESTMYSCSSRSRVESDEL 1320

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09818990/runat_26112002_093406_23337/app_query.fasta_1.1479
-DB=Published Applications_NA -QMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62
-TRANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09818990.ecgn.1.1.47 @runat_26112002_093406_23337
-NCPU=6 -ICPU=3 -NO_XLPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1510.5	21.9	1823	10	US-09-925-299-231 Sequence 231, App
2	1394	20.2	2059	10	US-09-764-853-218 Sequence 218, App
3	1133.5	16.5	702	10	US-09-764-853-397 Sequence 397, App
4	557.5	8.1	5926	10	US-09-969-708-302 Sequence 302, App

ALIGNMENTS

RESULT 1

US-09-925-299-231

; Sequence 231, Application US/09925299

; Patent No. US20020055627A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA102

; CURRENT APPLICATION NUMBER: US/09/925,299

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 231

; LENGTH: 1823

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (82)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc_feature

; LOCATION: (1593)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc_feature

5 557.5 8.1 5926 10 US-09-954-456-522 Sequence 522, App
6 527.5 8.1 5926 10 US-09-880-107-3371 Sequence 3371, App
7 1833 7.6 1833 10 US-09-880-192-32 Sequence 32, Appl
8 508.5 7.4 23907 9 US-10-077-130-6 Sequence 6, Appl
9 508.5 7.4 24120 9 US-10-077-130-4 Sequence 4, Appl
10 432.5 6.3 3192 10 US-09-925-300-502 Sequence 502, App
11 416.5 6.0 8883 10 US-09-905-129-1 Sequence 1, Appl
12 416.5 6.0 8883 10 US-09-905-129-5 Sequence 5, Appl
13 416.5 6.0 8883 10 US-09-905-129-7 Sequence 7, Appl
14 416.5 6.0 8883 10 US-09-991-630-1 Sequence 1, Appl
15 416.5 6.0 8883 10 US-09-991-630-5 Sequence 5, Appl
16 416.5 6.0 8883 10 US-09-991-630-7 Sequence 7, Appl
17 408.5 5.9 8180 10 US-09-905-129-8 Sequence 8, Appl
18 408.5 5.9 8180 10 US-09-991-630-8 Sequence 8, Appl
19 408.5 5.9 8262 10 US-09-905-129-6 Sequence 6, Appl
20 408.5 5.9 8262 10 US-09-991-630-6 Sequence 6, Appl
21 408 5.9 7770 10 US-09-991-630-23 Sequence 23, Appl
22 397 5.8 355 10 US-09-563-817-656 Sequence 656, App
23 396 5.7 9645 10 US-09-905-129-22 Sequence 22, Appl
24 396 5.7 9645 10 US-09-991-630-22 Sequence 22, Appl
25 390 5.7 443 10 US-09-960-352-13726 Sequence 13726, A
26 378 5.5 8106 9 US-10-077-130-1 Sequence 1, Appl
27 369 5.4 7893 9 US-10-077-130-3 Sequence 3, Appl
28 346 5.0 453 10 US-09-960-352-20 Sequence 20, Appl
29 337 4.9 431 10 US-09-960-352-8748 Sequence 8748, App
30 334 4.8 456 10 US-09-960-352-13256 Sequence 13256, A
31 327 4.7 2793 9 US-10-160-865-13 Sequence 13, Appl
32 326 4.7 2614 9 US-10-160-865-15 Sequence 15, Appl
33 323.5 4.7 4291 12 US-10-105-934-4 Sequence 4, Appl
34 322.5 4.7 468 10 US-09-960-352-932 Sequence 932, App
35 312.5 4.5 2259 12 US-10-105-934-6 Sequence 6, Appl
36 301 4.4 377 10 US-09-960-352-8061 Sequence 8061, App
37 296 4.3 407 10 US-09-983-965-508 Sequence 508, App
38 290 4.2 1083 10 US-09-764-853-373 Sequence 373, App
39 289 4.2 941 10 US-09-764-853-184 Sequence 184, App
40 285.5 4.1 417 10 US-09-960-352-933 Sequence 933, App
41 283 4.1 4682 10 US-09-783-066-8 Sequence 8, Appl
42 278.5 4.0 4978 12 US-10-044-090-89 Sequence 89, Appl
43 278 4.0 4092 12 US-10-124-557-51 Sequence 51, Appl
44 276.5 4.0 5510 10 US-09-880-107-1758 Sequence 1758, App
45 275 4.0 3420 12 US-10-124-557-103 Sequence 103, App

Alignment Scores:

Pred. No.: 1,1e-85 Length: 2059
Score: 1394.00 Matches: 261
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.62% Mismatches: 0
Query Match: 20.23% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-2 (1-1320) x US-09-764-853-218 (1-2059)

```
QY 1059 GlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAla 1078
Db 7 AGAGAAGAGACAAAGAGCCCTACAGGAACGCTTTTCCGACCACATTTCCCTCAGGCT 66
QY 1079 ProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGly 1098
Db 67 CCTGGGATATGTAGCTCATAGGGGGCGCTCTGCGCTGGACTGTAGGTAGTGGT 126
QY 1099 LeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSer 1118
Db 127 TTACCGCCCGGAGCTGACATGGCTACTCAATGGCCAACTGTGCTACCAGATGCCCTC 186
QY 1119 HisLysMetLeuValArgGluThrGlyValHisSerLeuLeuLeaspProLeuThrGln 1138
Db 187 CACAAGATGCTGTACGGGAGACCGGAGTCCACTCTGTCTCATTTGACCCACTCACTCAG 246
QY 1139 ArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSer 1158
Db 247 CGGACGACGGACCTATAAGTCATCGCTACCAACAAACCGGGCAGAAATCTTTTGT 306
QY 1159 LeuGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeu 1178
Db 307 CTGAGCTCTCTGTAGTACAAAGAGTGAAGAAAGCACCTGTGTATCTCTGGAGAACTA 366
QY 1179 GlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMet 1198
Db 367 CAGAACCTGCGGTGTCCGGAAGCCACCCCGTGAGACTGGAGTGCCGCGTATAGGCATG 426
QY 1199 ProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArg 1218
Db 427 CCCCCACTGTCTTACTCGAAGAAAGACAATGAGACCATCTCTTCATTCACCCAGAGAGG 486
QY 1219 IleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLys 1238
Db 487 ATCAGTATGCACAGGACACAAAGGTATGCTGCTGCTTCTCATTCACCCAGCAAGAAA 546
QY 1239 SerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThr 1258
Db 547 TCAGACGCTGGATGGTACAGTTGTACGCCAAGAAATGAAGCCGCAATCGTGTCTGCACCT 606
QY 1259 AlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnIleProProMetSerValArg 1278
Db 607 GCACGGCTGATATATACGCTCAGTGGCACCATCATGATCCACCGCCCATGTCTGTCCGG 666
QY 1279 ProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPhe 1298
Db 667 CCCAGTGGCAGTCGCTACGATCTCTCACAGTAAAGACTTGACATATTTTCTGCCTTT 726
QY 1299 SerSerMetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSerAsp 1318
Db 727 TCCTCCATGGAAAGCACGATGGTATTTCATGCTTCTCGGAGTGAGTGGAGAGTAT 786
QY 1319 GluLeu 1320
Db 787 GAACCT 792
```

RESULT 3

US-09-764-853-397
; Sequence 397, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ06

; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 397
; LENGTH: 702
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (587)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (661)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (670)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-853-397

Alignment Scores:
Pred. No.: 1,14e-68 Length: 702
Score: 1133.50 Matches: 226
Percent Similarity: 96.17% Conservative: 0
Best Local Similarity: 96.17% Mismatches: 9
Query Match: 16.45% Indels: 4
DB: 10 Gaps: 0

US-09-818-990B-2 (1-1320) x US-09-764-853-397 (1-702)

```
QY 1060 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 1079
Db 3 GAAAGAGACAAAGAGCCCTACAGGAACGCTTTTCCGACCACATTTCCCTGCAGGCTCCT 62
QY 1080 GlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeu 1099
Db 63 GGGGATATGTAGCTCATGAGGGCGCTCTGCGGTGGACTGTAAAGTGAGTGGTTTA 122
QY 1100 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 1119
Db 123 CG-CCCCGGAGCTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCCCTCCCAC 181
QY 1120 LysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArg 1139
Db 182 AGATGCTGTGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCACCTACTCAGCGC 241
QY 1140 AspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 1159
Db 242 GACGA-GGGACCTATAAGTGCATCGCTACCAACAAACCGGGCAGAATCTTTTAGTCTG 300
QY 1160 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGln 1179
Db 301 GAGCTCTCTGTAGTACAAAGAGGTGAAGAAGACACCTGTGTATCTCTGGAGAAACTACAG 360
QY 1180 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 1199
Db 361 AACTGCGGTGTTCGGAAGGCCACCCCGTGAGACTGGAGTGCCGCGTATAGGATGCC 420
QY 1200 ProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIle 1219
Db 421 CCACCTGTCTTACTTGAAGAAAGACAATGAGACCATCCCTTCCACCAGAGAGGATC 480
QY 1220 SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSer 1239
Db 481 AGTATGCACCGAGGACACAAAGGATGCTGCTTCTTCATTCAGCCAGCCAGCAAGAAATCA 540
QY 1240 AspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla 1259
Db 541 GACCTGGATGGTACAGTTGTACGCCAAGAAATGAAGCCGCGCATCGNGTCTGTGCACGTCC 600
QY 1260 ArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProMetSerValArgPro 1279
```

Db 601 AGGCTGGATATATACGCTCAGTGGCACCACCATCAGATCCACGCCCATGTC--TGCGGGGCC 658
Qy 1280 SerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIle 1294
|||||
Db 659 AGNGCCAGTCGNTAGGATCTCTACACT-AAAGGACTTGACATA 702
RESULT 4
US-09-969-708-302
; Sequence 302, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 302
; LENGTH: 5926
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-302

Alignment Scores:
Pred. No.: 1,23e-28 Length: 5926
Score: 557.50 Matches: 265
Percent Similarity: 33.71% Conservative: 124
Best Local Similarity: 22.96% Mismatches: 435
Query Match: 8.09% Indels: 330
DB: 10 Gaps: 42

US-09-818-990B-2 (1-1320) x US-09-969-708-302 (1-5926)

Qy 195 AsnSerSerPheSerAspLeuSerGluArgGluArgSerSerValProIlePro 214
:::|::|
Db 18 AGTTGACAGCAACTTCAGGAGCTTCCACGCGAGAGCTTCAGAGCGCTTTCCCTGTCCTCA 77
Qy 215 -----IleProAlaAspThrArgAspAsnGluValAsnHisAlaLeuGluGln 231
:::|::|
Db 78 CTGGCCCGAGTTGCCACAACAACACAGAGAGAGCGGTGACCATGGGG----- 125
Qy 232 GluAlaLysArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThr 251
:::|::|
Db 126 -----GATGTGAAGTGGTTCCTCGCTGCACACATT--TCCAAAACCTCC 167
Qy 252 ProGlySerSerProSerSerLeuTyrTyrGluProLeuGluGlnProProArgPhe 271
|||||
Db 168 CTCAGTGTGGATCCCTCAAGAGTT--GACTCCATGCCCTGCAGAGAGGCCCTGCTTTC 224
Qy 272 ThrGlnLysLeuArgSerArgGluValProGluGlyThrArgValGlnLeuAspCysIle 291
|||||
Db 225 ATTTTGGCCCCCTCGGAACCTTCGATCAAGAGAGGACCCCGCAAGTTCGAAGGGCGG 284
Qy 292 ValValGlyIleProProGlnValArgTrpTyrCysGluGlyLysGluLeuGluAsn 311
|||
Db 285 GTCGGGGTTTACCAGAGAGCCCGAGTGCATGGCACAGAAAACGGCAACCATC---ACC 341
Qy 312 SerProAspIleHisIleValGlnAlaGly-----AsnLeuHisSerLeuThrIleAla 329
|||
Db 342 AGCGGGGCCCTTCCTCGATGGATCGCGGCATCCGGGGAGCTTTCCAGCCTTGTGATTCAT 401
Qy 330 GluAlaPheGluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThr 349
|||||
Db 402 GCTGTCCATGAGGAGGACAGGGGAAGATATACCTGTGNAGGCCCAACCATGGCAGTGGTGCT 461

Qy 350 AspSerThrSerAlaGluIleTyrIleGluGlyValSerSerSerAspSerSerGluGlyAsp 369
:::|::|
Db 462 CGCCAGGTGACAGTGGAGTTGACAGTAGAAGGA---AGTTTGGCAAGCAGCTGGTGCAG 518
Qy 370 Pro-----AsnLysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerPro 387
|||
Db 519 CCTGTGTTTCCAAAACCTTAGGGGATAGATT----- 551
Qy 388 ProThrThrSerAlaValIleProProAlaValProGlnAlaGlnHisLeuValAlaGln 407
:::|::|
Db 552 ---TCAGCTTCAGCAGTG----- 569
Qy 408 ProArgValAlaThrIleGlnGlnCysGlnSerProThrAsnTyrLeuGlnGlyLeuAsp 427
|||
Db 570 ACCGTCCTAGCATCTGGGGGAGTGC----- 596
Qy 428 GlyLysProIleIleAlaAlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSer 447
|||
Db 597 -----CCACCAAGTTTGCTACCAAGCTGGCGCGAGTTGTGGTCAAA 638
Qy 448 GluGlyGlnLeuValPheGluCysArgValLysGlyAlaProSerProLysValGlu 467
|||||
Db 639 GAAGACAGATGGGACGATCTCTGCAAGATCACTGGCGGCCCAACCCGACGCTCACC 698
Qy 468 TrpTyrArgGluGlyThrLeuIleGluAspSerProAspPheArgIleLeuGlnLysLys 487
|||
Db 699 TGGCTCAAGGGAATGTTCCACTCAG-----CCGAGTGCCCGTGTG----- 740
Qy 488 ProArgSerMetAlaGluProGluGluIleCysThrLeuValIleAlaGluValPheAla 507
|||||
Db 741 -----TCTGTGTCTGAGAAGACGCGATCGAGTGTCTGGAATCCATGGAGTCAACCAA 794
Qy 508 GluAspSerGlyCysPheThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIle 527
:::|::|
Db 795 GATCAGCTGGGAGTGTACACGTGCTGCTGGTGAACGGTCCGGGGAAGGCCCTCATGTCA 854
Qy 528 AlaGlnLeuHisValArgGly----- 534
|||
Db 855 GCTCAACTTTCCATCCAAAGGTTTGGACAGTGCCTCAATAGTTCATTGTGAGAGAAACAAA 914
Qy 535 -----AsnGluAspLeu-----SerAsnAsn 541
|||
Db 915 GCCCAACTTCAGATGTCCAGAAAGAGGTGACCAATGTAATCTCAAGGAGTCCGAAGCTG 974
Qy 542 GlySerLeuHisSerAlaAsnSerThrAsnLeuAlaIleGluProGlnProSer 561
|||||
Db 975 GACAGCTCGGAGGTGAGCCCAAGCAAGCAACTGCTCCAGCCCCAGAGAGGTGGCTCC 1034
Qy 562 ProPro-----HisSerGluProProSerValGluGlnProProLys---ProLys 577
|||||
Db 1035 CCACCTCTGGGCTGCAACACAGCCAGCT-----CAGCCCCCAAGGAGTCCCAAG 1082
Qy 578 LeuGluGlyValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArg 597
|||||
Db 1083 CTGGAGTCAATGC-----AAGGACTCGCCCAAGACGGCC----- 1115
Qy 598 ValHisPheAsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyVal 617
|||
Db 1116 -----CCGACAGCCCCGCTCTTCAGAGAGCTTCAGCTCC----- 1151
Qy 618 ValThrThrArgGlnThrArgProAspSer***GlnGluArgPheAsnGlyGlnAlaThr 637
:::|::|
Db 1152 ATCACCTGAGGCGCGCAAGAGATTTCAG----- 1178
Qy 638 LysThrProGluProSerPhePro-----ValLysGluProProProValLeuAla 654
|||||
Db 1179 -----CCGGAACCAAGACAGCCAGCGCTGGGGTCTCTATCACCT----- 1217
Qy 655 LysProLysLeuAspSerThrGlnLeuGlnLeuHisAsnGlnValLeuLeuGluGln 674
|||
Db 1218 -----TCTGGA 1223
Qy 675 HisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe***MetThrVal 694


```
Db 1224 GAAGAGGAGGAGGCGCAGCTCCTCCCGTCCAGCCACCTTCCCCACGCGCAGCGCTGGC 1283
QY LeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLysAlaProSer 714
Db 1284 CTGGGAGCCAA-----GATGTTGTAGCAAGGCTGCTAACAGGAAATCCCCCATG 1334
QY 715 SerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnThrThrAla 734
Db 1335 GAGGGCCAGAGGATTCAGCATTCCTCCCAATTT----- 1367
QY 735 AlaThrValAlaProSerSerSerProValPheThrLeuSerSerThrProGlnThrIle 754
Db 1368 -----GAGAGCAAGCCCA----- 1382
QY 755 GlnArgThrValSerLysGlu-SerLeuLeuValSerHisProSerValGlnThrLysSe 774
Db 1383 -----AGCCAGGAGGTCAGGAAATCAAACTCTCAAGTTCAGATGTGAAGTT 1430
QY 774 rProGlyGlyLeuSerIleGln-AsnGluProLeuProGly-----ProThrGluP 792
Db 1431 TCGGGATTCAAAGCTGAAGTGGC-----CTGGTTCCTGGAAGCACCCTCCGTGAGG 1484
QY 792 roThrProProPheThrPheSerIleProSerGlyAsnGlnPhe-----GlnProA 810
Db 1485 AGACA-----GGAAGGCAGCATTTAGGTTTATGAA 1514
QY 810 rgCysValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheL 830
Db 1515 GATGCTGGCTCCCATTAACCT----- 1534
QY 830 euSerSerValLeuProSerLeuProAlaIleProPro-ThrAsnAlaMet***LeuPro 849
Db 1535 -----CTGCTGCTGAAAGCCCGGACC----- 1556
QY 850 ArgSerAlaProSerMetProSerGlnGlyLeuAlaLysAsnThrLysSerProGln 869
Db 1556 ----- 1556
QY 870 ProValAsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLys 889
Db 1557 -----AGGACAGTGGG----- 1568
QY 890 LysIleThrPheSerAspValArgProAsnGlnGlnGluThrLysIleSerSerPheGlu 909
Db 1569 -----ACATACAGCTGCATGCTTCAACGCCCAAGCCAGGCTGCTGTAGCTGG--- 1619
QY 910 GlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAsp 929
Db 1620 -----ACCTCCAAAGTGGAAAGGCTTGCCGTGATGGAG----- 1652
QY 930 AspGluIleGlnHisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLys 949
Db 1653 -----GTGCCCTCCTTCTCCAGT 1673
QY 950 ArgLeuLysHisPheArgValThrGluClySerProValThrPheThrCysLysIleVal 969
Db 1674 GTCTGAGGACTCGCTGTTATGAGCGCCAGGATTTGTGTCGACTGCTCCGTACGG 1733
QY 970 GlyIleProValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsn 989
Db 1734 GGGACCCAGTGCCCCGATCACTTGGCTGCTGAATGGCAGCCCATCCAGTAGCTCGC 1793
QY 990 GluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThr 1009
Db 1794 TCCACCTGCCAGGCC-----GGCGTGGCTGAGCTCCACATCCAGGATGCC 1838
QY 1010 ThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSer 1029
Db 1839 CTGCGGAGGACCATGACCTACACCTGCTAGCTGAGATGCTTGGGCGAGGTGCC 1898
QY 1030 CysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGly 1049
```

```
Db 1899 TGCAGCGCCTGGGTACCGTCCAT----- 1922
QY 1050 GlnSerHisArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArg 1069
Db 1923 -----GAAAGAAGAGTAGCAGGAAGTAGTACCTTCTGCTGTGGCTGCCAGCAAG 1976
QY 1070 PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeu 1089
Db 1977 CCCACTGCACCCATCTCTCTGCAGGGCTCTCTGATCTCAAGATCATGGATGGAAGCCAG 2036
QY 1090 CysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuAsn 1109
Db 2037 GTCACTATGACTGTCCAAGTGTGAGGAATCCACCCCTCAAGTCTGCTGCACAAAT 2096
QY 1110 GlyClnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis 1129
Db 2097 GGAATGAGATCCAAAGAGTCAGAGGACTTCCACTTTGAACAGAGAGAACT---CAGCAC 2153
QY 1130 SerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThr 1149
Db 2154 AGCTTTGGATCCAGGAAGTGTTCCTCCGAGGACACGGGACGTACACTCGGAGCCTGG 2213
QY 1150 AsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLys 1169
Db 2214 AACAGCGCTGGAGAGGTCGCGCACCCAGGCGCTGCTACGGGTACAAGAGCCTCACGATGGC 2273
QY 1170 LysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProVal 1189
Db 2274 ACCAGCGCTGGTTCATCAGTAAGCCTCGCTCAGTGACAGCCTCCCTGGGCCAGAGGTGC 2333
QY 1190 ArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAspAsn 1209
Db 2334 CTATCTCTCTGGCCCATAGCTGGTGACCCCTTCTACCGTGCACCTGGCTCAGAGATGGC 2393
QY 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAla 1229
Db 2394 AAAGCCCTCTGCAAAAGACACTGGCCACTTCGAGGTGCTTCAGATAGGAGCGTTCACCC 2453
QY 1230 CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLys 1249
Db 2454 ---CTGTTCTTAAAGAGGTGACGCCCTGGCATCGCCGCGCAGTATGATGATCCTGCTCAAG 2510
QY 1250 AsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis 1269
Db 2511 AACCGGTTGGCGAATCAGCTGCGAGGTGTCTACTGATGCTACAGACAGCTCTGCCAGA 2570
QY 1270 GlnIlePro-----ProMetSerValArgPro----- 1279
Db 2571 GCCTTCCAGGGGGAGGAGGAGCTGCCAGCTGCCAGGACCTCTGTGTGGAGGAGTTGGT 2630
QY 1280 -----SerGlySerArgTyrGlySerLeu 1287
Db 2631 GCTGATGGTGGTGTAGTGACCCGCTATGGGTCCCTG 2666

RESULT 5
US-09-954-456-522
; Sequence 522 Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
```



```
QY 850 ArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGln 869
Db 1556 -----1556
QY 870 ProValAsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLys 889
Db 1557 -----AGGCACAGTGGG---1568
QY 890 LysIleThrPheSerAspValArgProAsnGlnGlnCluThrLysIleSerSerPheGlu 909
Db 1569 -----ACATACAGCTGCACCTTCCACAGCCACAGCCAGGTGCTCTAGCTAGCTG---1619
QY 910 GlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAsp 929
Db 1620 -----ACCCCTCAAGTGGAAAGGCTTCCGCTGATGGAG-----1652
QY 930 AspGluIleGlnHisAspGluIleProThrGlyLysCysIleAlaProIlePheAspLys 949
Db 1653 -----GTGGCCGCCCTCTCTCCAGT 1673
QY 950 ArgLeuLysHisPheArgValThrGluGlySerProValThrPheThrCysLysIleVal 969
Db 1674 GTCTTGAAGACCTGCGCTGTTATTGAGGCCAGGATTTGTGTGCTAGCTCTCCGTACGG 1733
QY 970 GlyIleProValProLysValTyrrPheLysAspGlyLysGlnIleSerLysArgAsn 989
Db 1734 GGCACCCACAGTCCCGGATCACTTGGCTGCTGAATGGGAGCCCATCCAGTACGCTCGC 1793
QY 990 GluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThr 1009
Db 1794 TCCACCTCGCAGGCC-----GGCGTGGCTGAGCTCCACATCCAGATGCC 1838
QY 1010 ThrSerAspAspGlyAsnTyrrIleMetAlaAlaAsnProGlnGlyArgIleSer 1029
Db 1839 CTCCGGAGGACCATGTCACCTACCTGCTAGCTAGAGTACCTTGGGGCAGGTGCTC 1898
QY 1030 CysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGly 1049
Db 1899 TGCAGCGCTGGCTACCGTCCAT-----1922
QY 1050 GlnSerHisArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArg 1069
Db 1923 -----GAAAGAGAGTAGTACAGGAGAGTACCTTCTGCTGTGGCTCCACGCAAG 1976
QY 1070 PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeu 1089
Db 1977 CCACACTGCCACCATCTTCTGCGAGGGCTCTCTGATCTCAAAGTCAATGATGGAAGCCAG 2036
QY 1090 CysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsn 1109
Db 2037 GTCACATGACTGTCCAAAGTGTCCAGGAATCCACCCCTCAAGTCATCTGGCTGCACAA 2096
QY 1110 GlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis 1129
Db 2097 GGGAAATGAGATCCAAAGAGTCAGAGGACTTCCACTTTTCAACAGAGAGGAAGT---CAGCAC 2153
QY 1130 SerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrrLysCysIleAlaThr 1149
Db 2154 AGCCTTGGATCCAGGAAGTGTCCCGGAGGACAGGGGCGTACACCTCGGAGGCGCTGG 2213
QY 1150 AsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLys 1169
Db 2214 AACAGCGCTGGAGAGGTCGCCACCCAGCGCTGCTACGGTACAGAGCCCTCAGATGGC 2273
QY 1170 LysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProVal 1189
Db 2274 ACCAGCGCTGGTTCATCAGTAAGCCCTCGCTCAGTGACAGCCCTCCCTGGGCCAGAGTGC 2333
QY 1190 ArgLeuGluCysArgValIleGlyMetProProValPheTyrrTrpLysLysAspAsn 1209
Db 2334 CTCATCTCCTGCGCCATAGTGTGTGACCCCTTCTCCTACCGTGCACCTGGCTCAGAGATGC 2393
```

```
QY 1210 GluThrIleProCysThrArgGluArgIleSerMethHisGlnAspThrThrGlyTyra 1229
Db 2394 AAAGCCCTCTGCAAGACACTGGCCACTTCGAGGTGCTTCAGATGAGGAGCTTTCACC 2453
QY 1230 CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrrThrLeuSerAlaLys 1249
Db 2454 ---CTGTTTCTAAGAAAGTGCAGCCCTCGCATGCCGCCAGTATGAGATCTCTCAAG 2510
QY 1250 AsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrrAlaGlnTrpHisHis 1269
Db 2511 AACGGTGTGGCAATGTCAGTGTCCAGGTGTCTAGTGTACAGATGTACAGAACACAGCTCTGCCAGA 2570
QY 1270 GlnIlePro-----ProProMetSerValArgPro-----1279
Db 2571 GCCCTTCCACGGGGAGGAGGCTCGCCAGCTCGCAGGACCTCTGTGGTGGAGGATGGT 2630
QY 1280 -----SerGlySerArgTyrrGlySerLeu 1287
Db 2631 GCTGATGGTGTGTAGTAGTACCGCTATGGTCCCTG 2666
RESULT 7
US-09-880-192-32
; Sequence 32, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 3795510CT1
US-09-880-192-32
Alignment Scores:
Pred. No.: 5,75e-27 Length: 1833
Score: 523.50 Matches: 100
Percent Similarity: 66.08% Conservative: 50
Best Local Similarity: 44.05% Mismatches: 74
Query Match: 7.60% Indels: 3
DB: 10 Gaps: 2
US-09-818-990B-2 (1-1320) x US-09-880-192-32 (1-1833)
QY 1096 ValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuPro 1115
Db 673 GTGAGTGGACTGCCAGCTCTCGATGTGTGTCATGTATCTAAATGGAAGACAGATTCATCA 732
QY 1116 AspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuAspPro 1135
Db 733 GATGATTGTCACAAAATGATAGTGTCTGAGAAGGGTCTTCATCTCCTCATCTCTTGAAGTA 792
QY 1136 LeuThrGlnArgAspAlaGlyThrTyrrLysCysIleAlaThrAsnLysThrGlyGlnAsn 1155
Db 793 GTCAGAGCTTCAGATGACAGGGGCTTATGTCATGTGTGCCAAGATAGACGAGGAGGCC 852
QY 1156 SerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeu 1175
Db 853 ACCTTCACTGTGACAGCTGGATGCTCTTGCAAAAGACATAAAGACCAATGTTTATC 912
QY 1176 GluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgVal 1195
Db 913 TACAAACCACAGAGAAAAAAGTTTAGAGGGAGATTCAGTGAACACTAGATGCCAGATC 972
```

```
Qy 1196 IleGlyMetProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThr 1215
Db 973 TCGGCTATACCTCCACCAAGCTTTCTGGAAAGAAATATGAATGGTACATTCAAC 1032
Qy 1216 ArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnPro 1235
Db 1033 ACTGACCAATAAGCTTATATCAAGATAACACTGGAAGAGTTACTTTACTGATAAAGAT 1092
Qy 1236 AlalysLysSerAspAlaGlyTyrThrLeuSerAlaLysAsnGluAlaGlyIleVal 1255
Db 1093 GTAAACAAGAAAGATGCTGGGTGTATCTGTGCAGCAGTTAATGAAGCTGGAGTGACT 1152
Qy 1256 SerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisGlnIleProProPro--- 1274
Db 1153 ACATGTACACCAAGATTAGAGTTACGGCAGTCCAAACCAACTCTTCCAGCTCCTTAAG 1212
Qy 1275 ---MetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAsp 1293
Db 1213 CAGTTACGGGTTCGACCAACATTTCAGCAAAATATTTAGCAGCTTAATGGGAAAGTTTGAAT 1272
Qy 1294 IlePheSerAlaPheSerSerMetGluSerThrMetValTyrSerCysSerSerArgSer 1313
Db 1273 GTAAACAAGCTTTTAACCCA---GAAGAGAAATTTTCAGCGTTTGGCAGCTCAATCTCGA 1329
Qy 1314 ValValGluSerAspGluLeu 1320
Db 1330 CTCTATGAAAGTGAAGACTT 1350
RESULT 8
US-10-077-130-6
; Sequence 6, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Thereof
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6
Alignment Scores:
Pred. No.: 1.4e-24 Length: 23907
Score: 508.50 Matches: 302
Percent Similarity: 35.03% Conservative: 181
Best Local Similarity: 21.90% Mismatches: 487
Query Match: 7.38% Indels: 412
DB: 9 Gaps: 52
US-09-818-990b-2 (1-1320) x US-10-077-130-6 (1-23907)
Qy 122 ProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyrCys 141
Db 14984 CCGAGAGCTCTCTCTGGGTGAGTGAGTGACGATGCTTCCGGCGGCTGCCGTCGGGTGC 15043
Qy 142 SerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluLeu 161
Db 15044 ACCGGCTCTTCCGACCAACCAAGATCCGGCTGAAG----- 15076
Qy 162 SerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsnHis 181
Db 15077 -----TTTCAGATGAGGAGCTCTTCCTGAGTGACGAGGAGGCGCTGCAGAGCCAG 15127
```

```
Qy 182 LysSerLysLeuGluSerGlnAsn-----LysValMetGlnGluAsnSerSerSerPhe 199
Db 15128 AGGAGCCCGCGAGCTGGCAGACATACCAGAGATGAGCATTTTCATCTGCATCCGT-TTT 15186
Qy 200 SerAspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThr 219
Db 15187 GAGCGCTCACTAGGCGCCGCGAG-----ValAsnHisAlaLeuGluGlnGlnGluAla 233
Qy 220 ArgAspAsnGlu-----ValAsnHisAlaLeuGluGlnGlnGluAla 233
Db 15220 CGCTTCCAGGAGATGTTTCCACACTGGGCATTTGGGTGGAGATCAAGCTGGTGGACAG 15279
Qy 234 LysArgArgGluAlaGlnGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGly 253
Db 15280 GGGCTCGGAGGGTAGAGTGTGCATCAGCAA----- 15312
Qy 254 SerSerProSerSerSerLeuTyrTyrGluGluProLeu-----GlyGln 267
Db 15313 GAGACTCTGCTCCCTGGTGGTCCCTCCAGAGCATTTGCCAGCCTACTACTTCTGACGCT 15372
Qy 268 ProProArgPheThrGlnLysLeuArgSerArgGluValProGluGlyThrArgValGln 287
Db 15373 GCCCAGTGTTCCTGACTGAGTTGCAGAACCAAGATGTCAGATGGGTATCTCTGTGAGC 15432
Qy 288 LeuAspCysIleValValGlyIleProProGlnValArgTyrTyrCysGluGlyLys 307
Db 15433 TTTGACTCGGTGTGACAGGTTCAGCCCATGCCAGTGTGCGCTGTGTTCAAGATGGGAG 15492
Qy 308 GluLeuGluAsnSerProAspIleHisIleVal-----GlnAlaGlyAsnLeu 323
Db 15493 TTGTTGGAGGAG-----GATGATCACTACATGATTATGAAGACCAACAGGGTGGC 15543
Qy 324 HisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCysPheAla 343
Db 15544 CATCAGCTCATCATCAGACCGGTGTCACACAGATGGGCGTCTACCGCTGCTGTGCC 15603
Qy 344 SerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGlyValSerSer 363
Db 15604 GAGAACAGCATGGGTGCTCTCTCCACCAAGCTCAGCTCCGTGTGGAC---TTGACAAGC 15660
Qy 364 SerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLysProAsnGlu 383
Db 15661 ACAGACTATGAC----- 15672
Qy 384 ValSerSerProThrThrThrSerAlaValIleProProAlaValProGlnAlaGlnHis 403
Db 15673 -----ACTGCAGCA-----GATGCCAGGAGTCTCATCTCTAC 15705
Qy 404 LeuValAlaGlnProArgValAlaThrIleGlnGln-----CysGlnSerProThrAsn 421
Db 15706 TTCAGTGCCTCAAGGCTACCTGTCCAGCGGAGGAGGAGGAGGAGTCCACCACTGAT 15765
Qy 422 TyrLeuGlnGlyLeuAspGlyLysProIleAlaAlaProValPheThrLysMetLeu 441
Db 15766 -----GAGGGCCAG-----CTGCCCCAGGTGGTGGAGGAGCTG 15798
Qy 442 GlnAsnLeuSerAlaSerGluGly---GlnLeuValValPheGluCysArgValIleGly 460
Db 15799 AGAGCTTCCAGGTGGCGCCCTGGCCACCGCTGCCAGTTCACAGCTCAAGGTGAAAGGC 15858
Qy 461 AlaProSerProLysValGluTyrTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
Db 15859 TACCCTCTCCAGATATCTGTTCAAAAGATGGCCAGCCCTGACCGCATCTGCCAC 15918
Qy 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrIleu 500
Db 15919 ATCCGCATGACTGCGCAAGAAG-----ATCCTGCACACCTGT 15954
Qy 501 ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys 520
Db 15955 GAGATATCTCCGTCACCCGGGAGGACTCTGGCCAGTATGACGAGCTATATCAGCAATGCC 16014
Qy 521 TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn 540
```

Db 16015 ATGGGGTCTGCTACTCGTGTGCCCGGCTGCTGGTTTCAGGC----- 16056
QY 541 AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaAlaIleGluProGlnPro 560
Db 16057 -----CCT 16059
QY 561 SerProHisSerGluProProSer---ValGluGlnProProLysProLysLeuGlu 579
Db 16060 GATGAGCCAGAGAGAGCCGTGCATGATGTCATGAGCAGCTGGTGGCG----- 16110
QY 580 GlyValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArgValHis 599
Db 16111 -----CCCCGAATGCTGGAGAG----- 16128
QY 600 PheAsnLeuProGluAspLysGlySerGluAlaSerSerGluAlaGlyValValThr 619
Db 16129 TTCACCCCNAGAAAGTGAAGAAGCTCCAGCATCCTCTCTGTGAAGGTA----- 16182
QY 620 ThrArgGlnThrArgProAspSer***-----GlnGluArgPheAsnGly 634
Db 16183 -----GAAGGACGCCGGTGCCACCGTGCACCTGGCTCAGGAGGAGGCTGAGAGAGC 16236
QY 635 GlnAlaThrLysThrProGlu---ProSerPheProValLysGluProProValLeu 653
Db 16237 GTCTGTGGATTGGCCCTGCACACCGCGGCTACACCGTG----- 16275
QY 654 AlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGlu 673
Db 16276 -----GCCAGCTCTGCGCAGCAG---CACAGCTGTGCTCTGCTGGAG 16314
QY 674 -----GlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe 690
Db 16315 GTGGCGCGGAGCACCAG----- 16335
QY 691 **MetThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnVal 710
Db 16336 ACCTACACATGCATTGCGCAGCACAGCTGCGCGGCGAGCCCTCTGCTCCGCGCAGCTGCAC 16395
QY 711 LysAlaProSerSerGlnThrPheSerLeuAlaArgProLysThrPheProSerThr 730
Db 16396 GTCTCGGGCGCTGCTAGGTGGAGGAGCAGGAGAAAGTGAAGGAGCGCTGATTCCACT 16455
QY 731 AsnThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSer----- 748
Db 16456 TTCTCGAGGGGACCACACAGCCATCTCAGCACAGGGGTGGAACTCGCAGTTTGCT 16515
QY 748 ----- 748
Db 16516 GACCTTGGTGGCAGAGAGAAAGAGCCCTCTGGCTGCCAAGGAGGCCCTCGGCCACCTG 16575
QY 749 -----SerThrProGlnThrIleGlnArgThrValSerLysGluSerLeu 763
Db 16576 TCCTCGCTGAGTGGGCGCAGAGAGTCTCTGCAGAAATGACCTCCAGATCATCTGAG 16635
QY 764 LeuValSerHisProSerValGlnThrLys-----SerProGlyGlyLeuSerIleGln 781
Db 16636 ATGGTATCGCCCAAGATCAGCGAGCCCAAGCTGCAGTGGCCGGAGGTGACATGATGAG 16695
QY 782 AsnGluProLeuPro-----ProGlyProThrGluProThrPro-ProProPheTh 798
Db 16696 GACTCCAAAGACACCATCTGCATCCCCCGCATGGCCGATCAGCGGCATCTCCAGCATC 16755
QY 798 rPheSerIleProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSe 818
Db 16756 CAGGAGTCTTCTCCAGAGTCAG-----AGGACGGCGATGCCGAGCGCAGATC 16803
QY 818 rProThrSerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuPr 838
Db 16804 TTTGACATCT---ACGTGGTTCACCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGGATGCC 16860
QY 838 oAla-----IleProProThrAsnAlaMet** 847
| : : : : : | : : : : : |

Db 16861 ATCAGCGTGGGGAAGGCCAGTATGTGGAGTCTCTGGATGCAGCCCCACCACCTGCGTGG 16920
QY 847 *LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSe 867
Db 16921 CTTG-----TCCGACCAAGCCCAAGTCCAGCCCC--TCACGGCAGGGCTGGGTGTC 16973
QY 867 rPro----- 868
Db 16974 ACCAGCCTACCTGGACAGGAGGTCAAGCTGTCACTGAGTGGGGGCCCGCTGAGGCCCC 17033
QY 869 -----GlnProValAsnAspAsnIleArgGluThrLysAsnAlaValI 884
Db 17034 TGAGTTCCCTGGGAGGCTGTCTGAAGCAAGTACAAGCAAGGCTGAGCTGTGAT 17093
QY 884 eArgAspLeu----- 887
Db 17094 CCAGGAGCTCTGAGTTCTGACGAGGCTTCGTGGAGGAGCTGCAGTTCCTGACAGCCA 17153
QY 888 -----GlyLysLy 890
Db 17154 CCACCTGCACACCTGGAGCGCTGCCCCACGTGCCCATAGCCGTGCCGCCAAGAGGC 17213
QY 890 sIleThrPheSerAspValArgProAsnGlnGlnGluThrLysIleSerSerPheGluG 910
Db 17214 AGTCATCTTCCGCAATGTGGG-----GACATCGCGCGCTTCCACAG 17255
QY 910 nArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspAs 930
Db 17256 CAGCTTCTCCAGAG-----TTGCAGCAGTGGCAGCAGGACGACGACGT-GG 17302
QY 930 pGluIleGlnHisAspGlu-----IleProThrGlyLy 941
Db 17303 CCATGCTCTTCATCAAGAACACCGCGCCCTTTGAGCAGTACCTGGAGTTCTCTGGTGGGC 17362
QY 941 sCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGlu----- 958
Db 17363 GTGTGAGGCTGAGTGGTGTCTGTCAGCAGCGCCATCCAGGAGTTCTACAAGAAATACG 17422
QY 959 -----GlySerProValThrPheThrCysLysI 968
Db 17423 CGGAGGAGCGCTGTGGCAGGGAGCCCTCTCAGCCCCCGCCACCACCTCTCGAGCACT 17482
QY 968 eValGlyIle-ProValProLysValTrpPhe-----LysAspGlyLysG 984
Db 17483 ACCTGGAGCAGCCAGTGGAGGGGTGAGCGCTACACGAGCTTCTGAGAGGTTGATCC 17542
QY 984 InIleSerLysArgAsn---GluHisCysLysMetArgArgGluGlyAspGlyThrCys 1003
Db 17543 GCAACAAGGCGCGGAGACAGACAGAACTCCGCGCTGCTGGAGCAGGCTATGCCGTGTGT 17602
QY 1003 er-----LeuHisIleGluSerThrThrSerAspAsp- 1013
Db 17603 CTGCGCTGCCACAGCGCGCTGAGAAAGCTGCACGTGTCCCTCATGGAGAACTACCCAG 17662
QY 1014 -----AspGlyAsnThrThrIleMetAlaLa 1023
Db 17663 GCACCTTGAGGCCCTGGCGGAGCCCATCCGCCAGGCCACTTCTATCGTGTGGGAGGTG 17722
QY 1023 snProGlnGlyArgIleSerCysSerGlyHis-----L 1034
Db 17723 CACCGGGGCGCGCATGCCCTGGAAGGCCACAAACCGCTCAGCTGTCTCTTCGCCAACCC 17782
QY 1034 euMetValGlnSerLeuProIleArg---SerArgLeuThrSerAlaGlyGlnSerHisA 1053
Db 17783 ACCTGGTAATCTGCAAGCCCCCGCGGAGACTCCCGACCGATACCGCTCAGCTGTGTTC 17842
QY 1053 rgGlyArgSerArgVal----- 1058
Db 17843 GGAACATGATCAAGCTGAGCAGCATCGACCTGAAGCAGGAGTGGGGGGATGACCCCG 17902
QY 1059 -----GlnGluArgAsp-----LysGluProLeuGlnGluArg- 1069
Db 17903 CCTTCGAGGTGTGGCAGGAGGGGAGGACTCGGTGCGCAAGTACTGCTGTGACGAGCAGGA 17962
| : : : : : | : : : : : |

```
Qy 1069 ----- 1069
Db 17963 CAGCCATTATCAAGAGCTCGTGGGTGAAGAGATCTGTGGCATCCAGCAGCGTCTGGCCC 18022
Qy 1070 -----PheAeArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluG 1087
Db 18023 TGCCTGTGTGGCGGCCCGGACCTTGAAGAGAGAGCTGGCCGACTGTCACAGCCGAGCTGG 18082
Qy 1087 lYAtrGLeuCysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpL 1107
Db 18083 GTGACAGCTCAAGCTGCCTGCCGTGGTGAAGAGAGAGCTGGCCGACTGTCACAGCTGGT 18142
Qy 1107 euLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrG 1127
Db 18143 ACAAGATGGGAAGACAGTGCAGGTGGACCCACCACATCTCATTGAAGACCTGTATG 18202
Qy 1127 lYValHisSerLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTrpLysCysI 1147
Db 18203 GCTCGTGTGCACTCATCTGGACACGCTGACCGGTGTGACTCTGGCAGTACATGTGCT 18262
Qy 1147 leAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLys- 1166
Db 18263 TCGCGGCCGAGCCCGCTGGC-----AACTGCAGTACCCTGGGCAAGA 18304
Qy 1167 --GluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluG 1186
Db 18305 TCCTGGTGCAAGTCCACCACCGGTTCTGTGAACAGGTCCGGGCTCACCCCTTTGTGGAGG 18364
Qy 1186 lYHisProValArgLeuGluCysArgValIleGlyMetProProProValPheTrpL 1206
Db 18365 GAGAGGAGCGCCAGTTCACCTGCACCATCGAAGGCGCCGCTACCCGAGATCAGGTGGT 18424
Qy 1206 yLSAspAsnGluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrT 1226
Db 18425 ACAAGGAGCGGCCCTGCTGACCACTGGCAACAAGTTCACGACACTGAGTGAGCCTCGCA 18484
Qy 1226 hrGlyTrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTrpL 1246
Db 18485 GCGGCTGCTAGTGTGTGTATCCGGGCGGCGCAGCAAGGAGGACCTGGGCTCTACGAGT 18544
Qy 1246 euSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIle 1263
Db 18545 GTGAGCTGGTGAACGGCTGGGCTCCGCGGGGCTAGTGGGAGCTGCGCAT 18597
RESULT 9
US-10-077-130-4
; Sequence 4, Application US/10077130
; Patent No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE OF INVENTION: Members and Uses Therefor
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
; US-10-077-130-4
```

```
Alignment Scores:
Pred. No.: 1,41e-24 Length: 24120
Score: 508.50 Matches: 302
Percent Similarity: 35.03% Conservative: 181
Best Local Similarity: 21.90% Mismatches: 487
Query Match: 7.38% Indels: 412
DB: 9 Gaps: 52

US-09-818-990B-2 (1-1320) x US-10-077-130-4 (1-24120)
Qy 122 ProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTrpCys 141
Db 15055 CCGAGAGCTCTCTCTGGGGTGGAGTGGAGTGGCTTCCGCGCGGCTGCCGCTGGCTGC 15114
Qy 142 SerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGluLeu 161
Db 15115 ACCGGCTTCTCCGACCAAAAGTCCGGCTGAAG----- 15147
Qy 162 SerSerLysPheLysSerHisSerSerLysArgIleArgProAlaGlyLysAsnHis 181
Db 15148 -----TTTCAGATCAGGAGCTCTCTCTGAGTCAGAGAGGCGCTCCAGAGCCAG 15198
Qy 182 LysSerLysLeuGluSerGlnAsn-----LysValMetGlnGluAsnSerSerPhe 199
Db 15199 AGGAGCGCGGAGTGGCAGACATACCGGAAGATGAGCATTTTCATCTGCATCCGT-TTT 15257
Qy 200 SerAspLeuSerGluArgArgGluArgSerValProIleProIleProAlaAspThr 219
Db 15258 GAGCGCTCTACTAGGCGCCGCGAG-----GCAGTAACT 15290
Qy 220 ArgAspAsnGlu-----ValAsnHisAlaLeuGluGlnGlnGluAla 233
Db 15291 CGCTTCCAGGAGATGTTTCCACACTGGGCATTTGGGTGGAGATCAAGCTGGTGGACAG 15350
Qy 234 LysArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGly 253
Db 15351 GGGCTCGGAGGTGAGAGATGTGCATCAGCAA----- 15383
Qy 254 SerSerProSerSerLeuTyTrpGluGluProLeu-----GlyGln 267
Db 15384 GAGACTCTCCCTGCTGGTGGCTCCAGAGCCATTGCCAGCCACTGCCACTGCTTGTGAGCCT 15443
Qy 268 ProProArgPheThrGlnLysLeuArgSerArgGluValProGluGlyThrArgValGln 287
Db 15444 GCCCAGTGTTCCTGACTGAGTTGCAGAACCAAGAGTGCAGGATGGTATCTCTGTGAGC 15503
Qy 288 LeuAspCysIleValValGlyIleProProProGlnValArgTrpTrpCysGluGlyLys 307
Db 15504 TTTGACTCGCTGGTGCAGAGTGCAGCCATGCCAGTGTGGCTGGTTCAGAGATGGGAAG 15563
Qy 308 GluLeuGluAsnSerProAspIleHisIleVal-----GlnAlaGlyAsnLeu 323
Db 15564 TTGTTGGAGGAG-----GATGATCACTACATGATTATGAAGACCAACAGGGTGGC--- 15614
Qy 324 HisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTrpSerCysPheAla 343
Db 15615 CATCAGCTCATCATCACAGCGGTGGTCCAGACACATGGCGCTACCGCTGCTGGCC 15674
Qy 344 SerAsnIleTyTrpGlyThrAspSerThrSerAlaGluIleTyTrpIleGluGlyValSer 363
Db 15675 GAGAACAGCATGGGTGCTCTCCACCAAGGCTGAGCTCGCTGTGGAC---TTGACAAGC 15731
Qy 364 SerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLysProAsnGlu 383
Db 15732 ACAGCTATGAC----- 15743
Qy 384 ValSerSerProThrThrSerAlaValIleProProAlaValProGlnAlaGlnHis 403
Db 15744 -----ACTGCAGCA-----GATGCCACGGAGTCCCTCATCTCTAC 15776
Qy 404 LeuValAlaGlnProArgValAlaThrIleGlnGln-----CysGlnSerProThrAsn 421
Db 15776 ----- 15776
```


Db 15777 TTCAGTGCCCAAGGCTACCTGTCCACCGGGAGCAGGAGGAAACAGAGTCCACCACTGAT 15836
QY 422 TyrLeuGlnGlyLeuAspGlyLysProIleAlaAlaProValPheThrLysMetLeu 441
Db 15837 -----GAGGGCCAG-----CTGCCCCAGGTGGTGGAGAGCTG 15869
QY 442 GlnAsnLeuSerAlaSerGluGly----GlnLeuValValPheGluCysArgValLysGly 460
Db 15870 AGAGACCTCCAGTGGCCCTGGCCACACAGCTGGCCCAAGTTCACAGTCAAGGTGAAGGC 15929
QY 461 AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuLeuGluAspSerProAsp 480
Db 15930 TACCTGCTCCAGATATATCTGGTCAAGATGGCAGCCCTCAGCCGATCTGCCAC 15989
QY 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu 500
Db 15990 ATCCGGATGACTGGCAAGAG-----ATCTGCGCACCCCTG 16025
QY 501 ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys 520
Db 16026 GAGATCATCTCCGTACCCGGGAGGACTCTGGCCAGTATGCAGCCTATATACGAATGCC 16085
QY 521 TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn 540
Db 16086 ATGGGTGCTGCCTACTCGTCTGCCCGGCTGCTGGTTCGAGGC----- 16127
QY 541 AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnPro 560
Db 16128 -----CCCT 16130
QY 561 SerProHisSerGluProProSer---ValIleGlnProProLysProLysLeuGlu 579
Db 16131 GATGAGCCAGAGAAGCCCTGCATCAGATGTGCATGAGCAGCTGTGCGG----- 16181
QY 580 GlyValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArgValHis 599
Db 16182 -----CCCCAATGCTGGAGAG----- 16199
QY 600 PheAsnLeuProGluAspLysGlySerGluAlaSerSerGluAlaGlyValValThr 619
Db 16200 TTCACCCCAAGAAAGTGAAAGGCTCCAGCATCACCTCTCTGTGAAGGTA----- 16253
QY 620 ThrArgGlnThrArgProAspSer***-----GlnGluArgPheAsnGly 634
Db 16254 -----GAAGGACGCCGGTGCACCGTGCACCTGCTCAGGAGGAGGCTGAGAGAGC 16307
QY 635 GlnAlaThrLysThrProGlu---ProSerPheProValLysGluProProValLeu 653
Db 16308 GTCTGTGGATTGGCCCTGCACACCGGGCTACACCGTG----- 16346
QY 654 AlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGlu 673
Db 16347 -----GCCAGCTTGCACCACAG---CACAGCCTGGTCTCTGTGGAC 16385
QY 674 -----GlnHisGlnLeuGlnAsnProProProSerSerProLysGluPheProPhe 690
Db 16386 GTGGCCGCGCAGCACACAG-----GGC 16406
QY 691 ***MetThrValLeuAsnSerAsnAlaProProAlaValThrSer***LysGlnVal 710
Db 16407 ACCTACACATGTCATGTCACACACAGTGCAGCCGCTGCGGCGCCCTCTGCTCGCCAGCCTGCAC 16466
QY 711 LysAlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThr 730
Db 16467 GTCTCGGGCCCTGATAGGTGGAGGACGAGAGAACTGAAGGAGCGCTGATTTCACCT 16526
QY 731 AsnThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSer----- 748
Db 16527 TTCCTGCGGGGACCACACAGCCATCTCAGCACAGGGGTGGAAACTGCGAGTTTGCT 16586
QY 748 ----- 748
Db 16587 GACCTTGGTGGCAGAGGAAAGAGCCCTCTGGCTGCCAAGGAGGCCCTCGCGCCACCTG 16646

QY 749 -----SerThrProGlnThrIleGlnArgThrValSerLysGluSerLeu 763
Db 16647 TCCTCGCTGAGTGGCAGAGAGGAGTTCCTGCAGAACTGACCTCCAGATCACTGAG 16706
QY 764 LeuValSerHisProSerValGlnThrLys-----SerProGlyGlyLeuSerIleGln 781
Db 16707 ATGTATCGGCCCAAGATCAGCGAGGCAAGTGCAGGTGCGCCGAGGTGACAGTATGAG 16766
QY 782 AsnGluProLeuPro-----ProGlyProThrGluProThrPro-ProProPheTh 798
Db 16767 GACTCCAAGACACATCTGCATCCCCCGCCATGGCGGATCAGCGGCATCTCCAGCATC 16826
QY 798 rPheSerIleProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSe 818
Db 16827 CAGGAGTCTTCCTCAGATCAG-----AGGACGGCATGCCCGGCGAGATC 16874
QY 818 rProThrSerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuPr 838
Db 16875 TTTGACATCT---ACGTGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 16931
QY 838 oAla-----IleProProThrAsnAlaMet** 847
Db 16932 ATCAGCTGCGGAAAGGCCAGTATGTGGAGTCTCTGATGCAGCCACCATCTCGCTGG 16991
QY 847 *LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSe 867
Db 16992 CTGTG-----TCGCGACCAAGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTG 17044
QY 867 rPro----- 868
Db 17045 ACCAGCTACCTGGACAGGAGGCTCAAGCTGTCACTGTAGTGGGGGCCGTGAGGCC 17104
QY 869 -----GlnProValAsnAspAspAsnIleArgGluThrLysAsnAlaValI 884
Db 17105 TGAGTTCCTCGGGAGGCTGTGTGAAGACGAATACAAGCAAGGCTGCTGTGTGT 17164
QY 884 eArgAspLeu----- 887
Db 17165 CCAGGAGCTCTGAGTTCGTGAGCAGGCTTCGTGGAGAGCTGCAGTTCCTGACAGCCA 17224
QY 888 -----GlyLysLys 890
Db 17225 CCACCTGCAGCACCTGGAGGCGTCCGCCACGTGCCATAGCCGTGCCGCCAGAGGC 17284
QY 890 sIleThrPheSerAspValArgProAsnGlnGlnGluThrLysLysSerSerPheGlu 910
Db 17285 AGTCATCTTCGCAATGTGCGG-----GACATCGCGCGCTTCCACAG 17326
QY 910 nArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGluSerAspAs 930
Db 17327 CAGCTTCCTGCAGGAG-----TTGCAGCAGTGGCAGCAGCAGCAGCAGT-GG 17373
QY 930 pGluIleGlnHisAspGlu-----IleProThrGlyLys 941
Db 17374 CCATGCTCTTCATCAAGAACCAGCGCGCTTTGAGCAGTACCTGGAGTTCTCTGTGG 17433
QY 941 sCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGlu----- 958
Db 17434 GTGTGAGGCTGAGTGTGCTGTCAGCAGCGCCCTTGAGCAGGAGTTCTACAAGAAATACG 17493
QY 959 -----GlySerProValThrPheThrCysLysI 968
Db 17494 CGGAGGAGGCCCTTTCGCGAGGGGACCCCTCTCAGCCCCCCCCACCATCTTCGAGCACT 17553
QY 968 eValGlyIle-ProValProLysValTyrTrpPhe-----LysAspGlyLysG 984
Db 17554 ACCTGGAGCAGCCAGTGGAGCGGTGCAGCGCTACCAGGCCCTTCTGTAAGGAGTTGATCC 17613
QY 984 lnIleSerLysArgAsn---GluHisCysLysMetArgArgGluGluArgGlyThrCys 1003
Db 17614 GCAACAAGGGCGGGAACAGACAGAACTGCGGCTGCTGGAGCAGGCGCTATGCCCTGTGT 17673

```
Qy 1003 er-----LeuHisIleGluSerThrThrSerAsp- 1013
||
|||||:
Db 17674 CTGCCCTGCCACAGCGCTGAGAACAGCTGCAGTCTCCTCATGGAGAACCCAG 17733
|||||:
Qy 1014 -----AspGlyAsnTyrThrIleMetAlaAla 1023
|||||:
Db 17734 GCACCTGGAGGCGCTGGGCGAGGCCATCCGCCAGGCCACTTCATCGTGGGAGGTG 17793
|||||:
Qy 1023 snProGlnGlyArgIleSerCysSerGlyHis-----L 1034
|||
|||:
Db 17794 CACCGGGGCGCGATGCGCTGGAGGCCACACCCGTCACGTGTCCTCTTCGCGCAACC 17853
|||
Qy 1034 euMetValGlnSerLeuProIleArg-----SerArgLeuThrSerAlaGlyGlnSerHisA 1053
|||||
|||
Db 17854 ACCTGGTAATCTGAAGCCCGCGGAGAGACTCCCGCACCGATACCGTCAGCTACGTGTCC 17913
|||||
Qy 1053 rgGlyArgSerArgVal----- 1058
||
||:
Db 17914 GGAACATGATGAAGCTGAGCAGCATCGACCTGAACGACCGAGTGGAGGGGATGACCGCG 17973
|||
Qy 1059 -----GlnGluArgAsp-----LysGluProLeuGlnGluArg- 1069
|||||
|||
Db 17974 CTTTCGAGGTGTGCAGGAGCGGGAGGAGTCCGTGCGCAAGTACCTGCTGCAGGCACGGA 18033
|||
Qy 1069 ----- 1069
|||
Db 18034 CAGCCATTATCAAGAGCTCGTGGTGAAGGAGATCTGTGGCATCCAGCAGCGTCTGGCCC 18093
|||
Qy 1070 -----PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluG 1087
|||||
|||
Db 18094 TGCCTGTGTGGCGCCCGCGGACTTTGAAGAGGAGCTGGCGCATGCACAGCCGAGCTGG 18153
|||
Qy 1087 lyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpL 1107
|||
|||
Db 18154 GTGACAGCTCAAGCTGCCTCGCGCTGAGCGGCGCACACCCAGCCCTCATCAGCTGGT 18213
|||
Qy 1107 euLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrG 1127
|||||
|||
Db 18214 ACAAGATGGGAAGCAGCTGAGGTGGACCGCCACCCACCATCCTCATTTGAAGACCTGATG 18273
|||
Qy 1127 lyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysI 1147
|||||
|||
Db 18274 GCTCGTGTGCACCTCATCTCGACAGCCCTGACCGGTGTGGACTCTGGCCAGTACATGTGCT 18333
|||
Qy 1147 leAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLys- 1166
|||
|||
Db 18334 TCGCGGCCAGCGCCGCTGGC-----AACTGCAGTACCTCGGCGCTCACCCCTTTGTGGAGG 18375
|||
Qy 1167 --GluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluG 1186
|||||
|||
Db 18376 TCCGTGGTGCAAGTCCCAACAGCTTCGTGAACAAGGTCCGGGCTCACCCCTTTGTGGAGG 18435
|||
Qy 1186 lyHisProValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpL 1206
|||
|||
Db 18436 GAGAGGACGCCAGTTCCATCTGCACCATCGAAGCGCCCGCTACCCCGACATCAGGTGGT 18495
|||
Qy 1206 ysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrT 1236
|||||
|||
Db 18496 ACAGGACGGGGCCCTGCTGACCACTGGCAACAAGTTCACAGACACTGAGTGAGCTCGCA 18555
|||
Qy 1226 hrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrTrpL 1246
|||||
|||
Db 18556 GCGGCCCTGCTAGTGTGTGTATCGGGCGGCGCACCAAGGAGGACCTGGGGCTCTACAGGT 18615
|||
Qy 1246 euSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIle 1263
|||
|||
Db 18616 GTGAGCTGGTGAACCGCTGGGCTCCGCGGGGCTAGTGGCGAGCTGCGCATT 18668
|||
```

RESULT 10

US-09-925-300-502

; Sequence 502, Application US/09925300

; Patent No. US20020151681A1

; GENERAL INFORMATION:

```
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 502
; LENGTH: 3192
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3085)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-502
```

```
Alignment Scores:
Pred. No.: 1.59e-20 Length: 3192
Score: 432.50 Matches: 263
Percent Similarity: 34.74% Conservative: 141
Best Local Similarity: 22.61% Mismatches: 437
Query Match: 6.28% Indels: 323
DB: 10 Gaps: 46
```

US-09-818-990B-2 (1-1320) x US-09-925-300-502 (1-3192)

```
Qy 98 LysArgLeuSerProAspGlnMetLysHisSerPro-----AsnLeuSerPheGluPro 115
|||
|||
Db 87 AAGACCTATCGAAGACGACGTGAAGAGATCCAGCCGAGCAGAGTGGATTTCGTGCC 146
|||
Qy 116 AsnPheCysGlnAspAsnProArgSerProThrSerSerLysGluSerProGlnGluAla 135
|||
|||
Db 147 AACCTG-----CAGCGGCAA 161
|||
Qy 136 LysArgProGlnTyrCysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAla 155
|||||
|||
Db 162 GTGAAGCCAAAGACTGTCTCTGAG---GAAGAGAGAGAGAGTGCACAGCCCGCAGCGTC 218
|||
Qy 156 AspPheIleGluGluLeuSerSerLysPheLysSerHisSerLysArgIleArgPro 175
|||||
|||
Db 219 GATTTT-----CGCTCTGTCTGCCAAGAGGGGACTTCCAAACCCCGTGCCT 269
|||
Qy 176 Arg-----AlaCysLysAsnHisLysSerLysLeuGluSerGln 188
|||
|||
Db 270 GAGAAGGTGCCACCGCCAAACCTGCCACCCCGGATTTTCGCTCAGTCTGGGTGGCAAG 329
|||
Qy 189 AsnLysValMetGlnGluAsn---SerSerSerPheSerAspLeuSerGluArgArgGlu 207
|||||
|||
Db 330 AAGAAATATACAGCAGAGAATGGCAGCAGCAGTGCAGAGACCTGAATGCCAAGCAGCAGTG 389
|||
Qy 208 ArgSerSerValProIle-----ProIle----- 215
|||||
|||
Db 390 GAGAGTCCCAAGCCCTGAGCAATGCACAGCTTCAGGGCCCTTGAACCCGCTGGGCAAC 449
|||
Qy 216 -----ProAlaAspThr----- 219
|||
Db 450 GCCAAGCCTGCTGAGACCTTGAAGCCAATGGCAACGCCAAGCCTGCCGAGACCTGAAG 509
|||
Qy 220 -----ArgAspAsnGluValAsnHisAlaLeuGluGlnGlnAlaLys 234
|||
Db 510 CCCATGGCAATGCCAAGCCCTGATGAGAACCTGAAATCCCGCTAGCAAGAGAACTCAAG 569
|||
Qy 235 ArgArgGluAlaGluGlnAlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySer 254
|||
Db 570 AAAGACGTTAAGATGATGTGAACCTGCAAGAGAGAGCCCATGCAGGAGCAATAATGAA 629
|||
Qy 255 SerProSerSerLeuTyrTyrGluGluProLeuGlyGlnProProArgPheThrGlnLys 274
|||
```



```
Qy 878 -----GluThrLysAsnAlaValIleArgAspLeuGly- 888
Db 2577 AGTACATGGCAAGAGAAATGGCAGAAACGGCAATGCTGTG---AGAGCCATTGGAA 2633
Qy 889 -----LysLysIleThrPheSerA 895
Db 2634 GACTGTCTCTATGSCAATGATCTCAGGGCTCAGTGGCAGGAAATCTCTCAACAGGGTCA 2693
Qy 895 spValArgProAsnGlnGluThrLysIleSerPheGluGlnArgLeuMetAsnG 915
Db 2694 CAACAGCCCTCAATGCAGAA---AAACTAGATCTGAAGAAGATGTGCCAAGCTT 2750
Qy 915 luIleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisA 935
Db 2751 TCCTTGAGGCTGTGCTGAGGAAAGCCT-----CAT- 2782
Qy 935 spGluIleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheA 955
Db 2783 -----GTAAACCCCTATTCTCTAAGACCATTTCGCATTTAG 2819
Qy 955 rgValThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProL 975
Db 2820 AAGTTGTGGGGAAGTGTCTCTAGATTGACTGCAAGATTGAAGGATACCACGACCCCG 2879
Qy 975 ysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMeta 995
Db 2880 AGGTTGTCTGTTCAAAGATGACCAGTCAATC---AGSGAGTCCCGCCACTTCACAGATAG 2936
Qy 995 rgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspG 1015
Db 2937 ACTACGATGAGGAGCGGCAACGCTCTTTAAATTATTAGTGATGTTTGGCGGATGACGATG 2996
Qy 1015 lyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuM 1035
Db 2997 CCAAGTACACTCGAAGCTCTCAACAGCTCTTGAGGAAGCCACCTGCACACAGAGCTCA 3056
Qy 1035 etValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyA 1055
Db 3057 TTGTGGAAACGATG-----GAGGAAGGTGAAGNGA-AGGGGAA 3094
Qy 1055 rgSer---ArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProH 1074
Db 3095 GAGGAAGAAGTAGTAAACAAAGCCAGAGAAAGCAGTTTCTTAAGTCATATTAAAGGACT 3154
Qy 1074 lsPhe 1075
Db 3155 ATTTC 3159

RESULT 11
US-09-905-129-1
; Sequence 1, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Elnat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8883
; TYPE: DNA
```

```
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(8883)
; OTHER INFORMATION: 'n' can be any nucleotide 'a', 'c', 'g' or 't'.
US-09-905-129-1
Alignment Scores:
Pred. No.: 6,73e-19 Length: 8883
Score: 416.50 Matches: 354
Percent Similarity: 30.73% Conservative: 216
Best Local Similarity: 19.08% Mismatches: 552
Query Match: 6.04% Indels: 733
DB: 1.0 Gaps: 83
US-09-818-990B-2 (1-1320) x US-09-905-129-1 (1-8883)
Qy 2 GlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu 18
Db 1460 CAGGAGGACAATGGATCTGCCTCCACCTCAAGATTTTCATAGAACCCCTTTGGCTCC 1519
Qy 19 SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArgAla 35
Db 1520 TTGTCTTTGAACATGACANANNNTCTGGAAATAAGGCCGACATGGTGTGTAGTATCCAA 1579
Qy 36 GluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 55
Db 1580 AAGCCATCAAGGACA-----TCACCACTGCATTCACCTGAA-----1615
Qy 56 GlyGlnAspAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSer 75
Db 1616 ---GAAATGCTACATCATCTAAATCGCTCATTT-----TCC 1651
Qy 76 ValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGln 95
Db 1652 ACAATCTGTG---TGCAGTGTAGATTATAATCACATCCAGCCAGTGTGGCACTTCTG 1708
Qy 96 AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluPro 115
Db 1709 GCTTATACAGTGACTCTCTCTGATAGAAAGGAGCCAGCTTACC-----1759
Qy 116 AsnPheCysGlnAspAsnProArgSerProThrSerLysGluSerProGlnGluAla 135
Db 1760 -----GAGACTCTTCACTGTCTTCTAGATATAAACAGGTGGCT 1798
Qy 136 LysArgProGlnTyrCysSerGluThr-----144
Db 1799 CTTAGGCTGAAGACATTTTACCAGCATAGAGCTGATGTCCAGCAGACGCCCTTTTGG 1858
Qy 145 ---GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluLeu 161
Db 1859 TTCCAACAAGAAAAAATGTCTTCAGCTGAACAGAACTGCCACCACACTTAGCACATTA 1918
Qy 162 SerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsnHis 181
Db 1919 CAGATCCAGTTTTCACCTGATGCTCAAAATCGCTTTA---CCAAGGGCG---GAGATGAGA 1972
Qy 182 LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSerAsp 201
Db 1973 GCGGAGAGACTCAAAATGGACCATGATCTGTATGATGAACAAT-----2014
Qy 202 LeuSerGluArgArgGluArgSer-----SerValPro 212
Db 2015 -----CCCAAACTGGAACGCACTGCTCTGTTGGCGSCACTATTGCCCTCAGCTGTCCA 2068
Qy 213 -----IlePro 214
Db 2069 GGCAAGCGCACCTTCCACCTCACTTGGATGGCTTCTAGCTGATGGAGTAAAGTGAGA 2128
Qy 215 IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln 230
Db 2129 GCCCTTACGTTGCGGAGGATGGCGGAATCTTAATAGACAAAAATGGGAAGTTGGAACGTG 2188
```

Qy	231	GlnGluAlaLysArgGluAla	-----GluGlnAlaAlaSerGluAlaAlaGly	247
Db	2189	CAGATGCTCAGACGCTTTGATGCAGGCTTTTACCACCTGCATAAGCACCAATGATGCAGAT		2248
Qy	248	GlyAspThrThrProGlySerSerProSerLeuTyrTyrGlu	-----Glu	263
Db	2249	CGGATGTT		
Qy	264	ProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGly	-----CTCACATACAGGATTAACGTGCTGGTAGAG	2284
Db	2285	CCCTATGGAGAAACACACATGACAGTGGAGTC		2338
Qy	284	ThrArgValGlnLeuAspCysIleValValGlyIleProProProGlnValArgTyrTyr	-----CAGCACACAGTGGTTACGGGT	303
Db	2339	GAGACGCTCAGACCTCCATGCCTTTCCACGGGTGTTCCAGATGCTTCTATTAGCTGATT		2398
Qy	304	CysGluGlyLysGluLeuGlnAsnSerPro	-----AspIleHisIleValGlnAlaGly	321
Db	2399	CTTCCAGGGAACACTGTGTCTCAGCCCTCAGAGACAGCAAAATCTTAAACAATGGG		2458
Qy	322	AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys	-----CCTTAAAGATCAAGGTCATTACCAATGT	341
Db	2459	ACCTTAAAGATATTACAGGTTACG		2509
Qy	342	PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIle	-----	358
Db	2510	GTGGCTCCCAACCCATCAGGCGGCGACTTTCCAGTTTTAAAGTTTTCAGTTCAAAAGAAA		2569
Qy	359	--GluGlyValSerSerSerSerGlu	-----GlyAspProAsn	371
Db	2570	GGCCAAAGATGTTGAGCATGACAGGAGGAGCGAGTGGATCTGGAGAACCCCAAC		2629
Qy	372	LysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerProProThrThrSer	-----	391
Db	2630	TCCAGTGTTCCTTAAAGCAGCCAGCATCTTTGAAACTCTCTGCATCAGCTTTCACAGG		2689
Qy	392	AlaValIleProAlaValPro	-----	399
Db	2690	TCAGAGGCTGGAACAAAGTCTCCGGTGTTACATAGAGAACAACATAGAGACTTAATA		2749
Qy	399		-----	399
Db	2750	CATCGCGCGCTGGGATTCACGCTCCGGCGATTTCAGGAGCATAGGAGCAGCTCCCT	-----	2809
Qy	400	--GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGln	-----	417
Db	2810	CTCTCTCTCGGAGATTGACCGCAACGCTGGCGAGCATCTTAGAAAAAGCCAAAAG		2869
Qy	418	--SerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIle	-----IleAlaAlaPro	435
Db	2870	AATTCTGTGCAAAAAGCAAGAAATATACACAGATAAGCCAGTGCCTGCTGCTTCCC		2929
Qy	436	ValPheThrLysMetLeuGlnAsnLeuSerAlaSer	-----GluGlyGln	450
Db	2930	CTCTGTGAACCTACTGACGAGGAAAAGGATGCTCTGGCATGATTCCTCCAGATGAAGAA		2989
Qy	451	LeuValValPheGluCysArgValLysGlyAlaPro	-----SerProLysValIle	467
Db	2990	TTCATGTTCTGAAACTAAGGCTTCTGGTGTCCCGAGGAGGTCAACCACTGCTGACTCT		3049
Qy	468		-----TrpTyrArgGluGlyThrLeuIle	475
Db	3050	GGACCAGTAATCATGTGTTTATGACGAGTATAGTCTTGGCACAGAAAGTCTCAACTGTG		3109
Qy	476		-----GluaspSerProAspPheArg	482
Db	3110	AATCCAAACACTACAATCTGAGCACCCTTCCTGTGATTTCAAAATATTATTAGTGTAAACAAC		3169
Qy	483		-----IleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThr	499
Db	3170	GGTACACTGTGCAAAAGATATGAACCCATCCATAGCAGCAAAATAGAAGATACAAACC		3229
Qy	500		-----LeuValIleAlaGluValPheAlaGlu	510

Db	3230	AACCAAAACCAATCATATATCTTTCCATCAGTAGCTGAAATTCGAGATTCTGCTCAGGCA	3289
QY	510	510
Db	3290	GGAAGAGCATCTCCCAAAAGTGCACACCCTGTAAACAGGGGGAAACATGGCTACCTATGGC	3349
QY	511	-----GlyCysPheThrCysThrAlaSer- 	518
Db	3350	CATACCAACACATATAGTAGCTTTACCAGCAAGCCAGTACAGCTTGTGCAGCCAATAAAT	3409
QY	519	-----AsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGlu	536
Db	3410	CCACAGAAAAGTTATGGACCTCAGATACCTATTACAGGAGTGCAGCAGACCTAGCAGTAGT	3469
QY	537	AspLeuSerAsnAsnGly-----SerLeuHisSerAlaAsnSerThr 	550
Db	3470	GACATCTCTTCACACTACTGCAGACCCTAGTCTTCAGTCACCCCTTCAGGTTCCACAC	3529
QY	551	ThrAsnLeuAlaIleGluProGlnPro- 	560
Db	3530	ACCAGCTGCTCGTCTTATTTCACATTCTCTAGAAAACAATACAGGTAACTTCCTCCCTTG	3589
QY	561	-----SerProProHis 	564
Db	3590	TCCAGGCCTTTGGAGAGAGAGAGACAATTTGGAGCAGAGGAGATTAAAAACCCACAT	3649
QY	565	SerGluPro----- 	567
Db	3650	AGAACCCGAGTCTCCGACGGCATAGACAGGACTGTGAGGCCAGCAATCAAGGGAGCT	3709
QY	568	-----ProSerValcGlnPro----- 	573
Db	3710	GCTAACAAAATGTGACCAAGTTCACGCCACAGACTACCTGGGATGTGCCACACATGT	3769
QY	573	----- 	573
Db	3770	CCTTCGCGAGGGGCTCACAGTGGTCTGTGCAGCACTGTGAGTTCACAGTTTCATCCCAAC	3839
QY	574	-----ProLysProLysLeuGluGlyValLeuVal----- 	583
Db	3830	AGTGCCTCCCAAACTAATAATGTGGGGTCTATAGCAGAGAGTCTACCACTGTGGTC	3889
QY	583	----- 	583
Db	3890	AAGAAACCAGCTGTACTATTTAAGGACAAACAATGTAGATATTGAGATAATACAACC	3949
QY	584	-----AsnHisAsnGluProArgSerSerSerArgIle 	594
Db	3950	ACTCAAAATATTCGGAGGGGAAAGTAACACAGCTGATTCTCAGGAAGCAAGCATGACT	4009
QY	595	GlyLeuArgValHisPheAsnLeu-----ProGluAspAspLysSerGluAla 	611
Db	4010	TCTGCTCCACATCTGTATCCCTGGGGAATCTCTGTAGACAATAGTGGT-----	4060
QY	612	SerSerGluAlaGlyValValThrThrArgGlnThrArgProAspSer***GlnGluArg 	631
Db	4061	-----CACCTGACATGCCTGGGACCATCCAAACTGGGAAGATTCAGTGGAA-----	4108
QY	632	PheAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProPro-----	650
Db	4109	-----ACAACACCACCTTCCCAAG-----CCGCTCAGCACACCTCAATA 	4147
QY	651	ProValLeuAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeu-----HisAsn 	668
Db	4148	CCAAACAGCAAAATTTCTCAAAGAGGAAACCTCCCTTCACCAAGATCTTTGTAATAAC	4207
QY	669	Gln-----ValLeuLeuGluGlnHisGlnLeu-----GlnAsnProPro 	681
Db	4208	CAGAGAAGGAGGGAGTGTAAAGAATCCATATCAATTCGGTTTACAAAGAACCCAGCC	4267
QY	682	ProSerSerProLys-----GluPheProPhe***Met 	692

[illegible]

Qy	916	IleGluPheArgLeuGluArgThrProValAspSerAspGluIleGlnHisAsp	935
Db	5261	AAAGGGCAGCAGTAAGCATGTCCCC	5293
Qy	936	GluIleProThrGlyLysCysIleAla-	944
Db	5294	AGCTTTCAGAGCCAGCACTCATGCTCTACACTGGAATACACAGAAGCAGTGCAGAAAG	5353
Qy	945	ProIlePheAspLysArg-	950
Db	5354	AGTGTGTTTGTATAGAAACCTGGTCAAAACCCCAACTCTCCAAACATCTGCCTTACGTCTCT	5413
Qy	951	-----LeuLysHisPheArgValThrGluGlySerProValThrPheThr	965
Db	5414	CTACCTAAGACTCTATTGAAAAAGCCAAAGAAATAATTGGAGAAAGCGTGCAAGCTTTACA	5473
Qy	966	-----CysLysIleValGlyIleProValProLys	975
Db	5474	GTTCCAGCTAATTCAGACGTTTTTCTCTGTGAGGCTGTGGAGACCACTGCCATC	5533
Qy	976	ValTyrTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGlnHisCysLys	993
Db	5534	ATCCACTGGACAGAGTTTCATCAGGANTTGAATATATCCCAAGGGACACACAAAAAGCGG	5593
Qy	994	MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp	1013
Db	5594	TTCACAGTGTTCCCAATGGACCC-----TTGTCCATCCAGGGGTGAGTATTCAGCAG	5647
Qy	1014	AspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis	1033
Db	5648	CGTGGACAGTACCTGTGCTCTGCATTTAATCCACTGGGCGTAGACCAATTTTCATGTCTCT	5707
Qy	1034	LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyClnSerHisArg	1053
Db	5708	TTGTCTGTGGTTTTTACC	5728
Qy	1054	GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArgPro	1073
Db	5729	-----GCAAGGATTTGGACACACATGTCAAG-----	5755
Qy	1074	HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAsp	1093
Db	5756	-----GAGATCCAGTTCATCTTTGGAAGTACTGTGGAACATAAG	5794
Qy	1094	CysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyClnProVal	1113
Db	5795	TGCAGAGTGGAGGTATGCCAGGCCCTACGGTTCTCTGGATACTTGCAACCAACGCGT	5854
Qy	1114	LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer	1130
Db	5855	GTCTCAGAAACGGCCAAAGGAAGCAAGGCTGGGTAAACCTGATGGA-----ACA	5908
Qy	1131	LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsn	1150
Db	5909	TTGATCATCTATATCATGAGCTTTTATGATCGGTGTTTTACAACTGTGGCGACGAGCA	5968
Qy	1151	LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys	1170
Db	5969	CCATCTGCCAGGATTCACCTGTTGGTTAAGATACAGTCATCACA-----GCT	6016
Qy	1171	AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro	1188
Db	6017	CCCCCTGTCTATTAGACAAAAGAGGCAAGCCATCGTTGGGGTT---TTAGTGGGAAGT	6073
Qy	1189	ValArgGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAsp	1208
Db	6074	TTGAACTGCCCTGCACCTGCAAAAGGAAGCACTCCCGACCTAGTGTCTACCTGGGTCTTTAT	6133
Qy	1209	AsnGluThr-----IleProCysThrArgGluArgIleSerMethIleGlnAsp	1224
Db	6134	GATGGGACTGAACATAAACCATTTGCAGTGTGACTTCATCCAGATTTTTCTGTATGCCAAAT	6193

Qy	994	MetArgArgGluGlyaspGlyThrCysSerLeuHisIleGluSerThrThrSerAspasp	101
Db	5594	TTCCACGTGCTTCCCAATGACG-----TTTGTCCATCCAGAGGTCAGTATTCCAGGAC	5647
Qy	1014	AspGlyAsnTyThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis	1033
Db	5648	CGTGGCAGTACCTGCTCTGCTGCTTTAATCCACTGGGCGTAGACCATTTTCATGTCCTCT	5707
Qy	1034	LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArg	1053
Db	5708	TTGTCCTGCTGTTTACCG-----	5728
Qy	1054	GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgPro	1073
Db	5729	-----GCAAGGATTTTGGAGACACATGTCAAG-----	5755
Qy	1074	HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuLeuAsp	1093
Db	5756	-----GAGATCACAGTTCATTTGGAGAGTACTGTGGAACCTAAAG	5794
Qy	1094	CysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVal	1113
Db	5795	TGCAGATGGAGGTATGCGCGGCGCTACGGTTTCTCGTATCTTGCAACCAACGCTG	5854
Qy	1114	LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer	1130
Db	5855	GTCTCAGAAACGGCCAGGCAAGGAGTCTGGGTAACACCTGATGGA-----ACA	5908
Qy	1131	LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrlLysCysIleAlaThrAsn	1150
Db	5909	TTGATCATCTATATCTGAGTCTTTATGATCGGTGTTTTACAAAGTGTGTGCCAGCAAC	5968
Qy	1151	LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys	1170
Db	5969	CCATCTGCCAGGATTCATCTGTTGGTTAGATPACAAGTCAATCACA-----GCT	6016
Qy	1171	AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro	1188
Db	6017	CCCCCTGTCTATTATAGACAAAAGAGCAAGCCATCGTTGGGTT---TTAGTGTGAAGT	6073
Qy	1189	ValArgLeuGluCysArgValIleGlyMetProProProValPheTyrlTrpLysLysasp	1208
Db	6074	TTGAACATGCCCTGCACCTGCAAGGAAGCAATCCCGACCTAGTGTTCACGTGGCTTTAT	6133
Qy	1209	AsnGluThr-----IleProCysThrArgGluArgIleSerMetHisGlnAsp	1224
Db	6134	GATGGGACTGNACTAAACCAATTCAGTTCATTCACAGATTTTCTTGATTCCAAT	6193
Qy	1225	ThrThrGlyTyrlAcysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrl	1244
Db	6194	GGAAC-----CTGTATATAAGAAGCATCGCTCCTTCAGTGAGGGGCACCTTAT	6241
Qy	1245	ThrLeuSerAlaLysAsnGluAlaGly-----	1253
Db	6242	GAGTGCATTTGCCACACCTCCTCAGGCTCAGAGAGAAGGTAGTGATCTTACTCTGGAA	6301
Qy	1253	-----	1255
Db	6302	GAGGAGAGACATCCCGCAGATAGAACTGCCCTCTCAGAAATGGACTGAGTGAATTG	6366
Qy	1254	-----IleValSerCysThrAlaArgLeuAsp-----	1266
Db	6362	GGTGAGAAATTACTACTGAACTGCTCAGCTACTCTGGGATCCAAAGCCTAGATAATCTGG	6422
Qy	1263	-----IleTyrlAlaGlnTrpHisHisGlnIleProProMetSer	1276
Db	6422	AGGCTGCCATCCAAAGGTGTCACGACCATGG---CACAGAATGGCAGCGCAATCCAC	6478
Qy	1277	ValArgProSerGlySerArg---TyrGlySerLeuThrSerLys	1290
Db	6479	CTTACCCAAATGATCTTGGTGGTTGGTGGCTGAGTCAGCGCAAA	6523

US-09-905-129-7

; Sequence 7, Application US/09905129
; Patent No. US20020137705A1

GENERAL INFORMATION:

; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7

LENGTH: 8883

TYPE: DNA

ORGANISM: Rattus sp.

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(8916)

; OTHER INFORMATION: 'n' can be any nucleotide 'a', 'c', 'g' or 't'.

US-09-905-129-7

Alignment Scores:

Pred. No.:	6.73e-19	Length:	8883
Score:	416.50	Matches:	354
Percent Similarity:	30.73%	Conservative:	216
Best Local Similarity:	19.08%	Mismatches:	552
Query Match:	6.04%	Indels:	733
DB:	10	Gaps:	83

US-09-818-990b-2 (1-1320) x US-09-905-129-7 (1-8883)

Qy	2	GlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu	18
Db	1460	CAGGAGGACAATGATGTCCTCCACCTCAAGATTCATAGAACCCCTTTGGGTCC	1519
Qy	19	SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArg	35
Db	1520	TTGCTTTGAACATCAGACANNTTCTGGAATAAGGCCGACATGGCTGTAGTATCCAA	1579
Qy	36	GluProSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlygly	55
Db	1580	AAGCATCAAGGACA-----TCACCACTGCATTCACTGAA-----	1615
Qy	56	GlyGlnAspAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSer	75
Db	1616	---GAAATGACTACATCATCTAATCGTCATTT-----TCC	1651
Qy	76	ValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGln	95
Db	1652	ACAATCTTGTG---TGCAGTGTAGATTATATACATCCAGCCAGTGTGCCAATCTCTG	1708
Qy	96	AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluPro	115
Db	1709	GCTTTATACAGTACTCTCCTCTGATACTAGAAAGGAGGCCACGCTTACC-----	1759
Qy	116	AsnPheCysGlnAspAsnProArgSerProThrSerSerLysGluSerProGlnGluAla	135
Db	1760	-----GAGACTCCTTCACTGCTCTTAGATATAAACAGGTGGCT	1798
Qy	136	LysArgProGlnTyrCysSerGluThr-----	144
Db	1799	CTTAGGCTGAAGACATTTTACCAGCATAGAGCTGTATGTCAGAGCAGACCCCTTTTGG	1858
Qy	145	---GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluLeu	161

Db	1859	TTCCAACAAGAAAAAATGTCTTGAGCTGAACAGAACTGCCACACACTTAGCACATTA	1918
Qy	162	SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis	181
Db	1919	CAGATCCAGTTTCCACTGATGCTCAATCGCTTTA---CCAAGGGC---GAGATGAGA	1972
Qy	182	LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSerAsp	201
Db	1973	GCGAGAGACTCAATGGACCATGATCTGATGATGAACAT-----	2014
Qy	202	LeuSerGluArgArgGluArgSer-----SerValPro	212
Db	2015	-----CCCAAACTGGAACGCACTGCTCTGTTGGCGCACTATTGCGCTGAGCTCCA	2068
Qy	213	-----IlePro	214
Db	2069	GGCAAGCGCACCTTCACCTCAGTTGGAATGGCTTAGCTGATGGAGTAAAGTGA	2128
Qy	215	IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln	230
Db	2129	GCCCTTACGTTAGCGAGGATGGCGAATCTAATAGACAAATGGGAAGTTGGAACGT	2188
Qy	231	GlnGluAlaLysArgArgGluAla-----GluGlnAlaLysSerGluAlaAlaGly	247
Db	2189	CAGATGGCTGACAGCTTTGATGCAGGCTTTTACCACCTGCATAAGCACCAATGATGCAGAT	2248
Qy	248	GlyAspThrThrProGlySerSerProSerSerLeuTyrTyrGlu-----Glu	263
Db	2249	GCGGATGTT-----CTCACATACAGGATAACTGTGTTAGTAGAG	2284
Qy	264	ProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGly	283
Db	2285	CCCTATGGAGAAACACACATGACAGTGGAGTC-----CAGCACAGAGTGGTTACGGGT	2338
Qy	284	ThrArgValGlnLeuAspCysIleValValGlyIleProProGlnValArgTyrTyr	303
Db	2339	GAGACGCTCGACCTTCCATGCTTCCACGGGTCTCCAGATGCTTCTATTAGCTGGATT	2398
Qy	304	CysGluGlyLysGluLeuGluAsnSerPro-----AspIleHisIleValGlnAlaGly	321
Db	2399	CTTCCAGGGAACACTGTGTTCTCTCAGCCATCAAGACAGAGCAAAATCTTAAACAATGGG	2458
Qy	322	AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys	341
Db	2459	ACCTTAAGATATTACAGTTACG-----CCAAAAGATCAAGGTCAATCAATGT	2509
Qy	342	PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIle-----	358
Db	2510	GTGGCTGCCAACCCATCAGGGCGGCACTTTTCCAGTTTAAAGTTTCAAGTTCAAAAGAAA	2569
Qy	359	---GluGlyValSerSerSerAspSerGlu-----GlyAspProAsn	371
Db	2570	GGCCAAAGGATGGTTGACATGACAGGAGGACGAGTGGATCTGCACTTGGAGAACCAAC	2629
Qy	372	LysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerProThrThrSer	391
Db	2630	TCCAGTGTTCCTTATAGCAGCCAGCATCTTTGAACACTCTCTGCATCAGCTTACAGGG	2689
Qy	392	AlaValIleProProAlaValPro-----	399
Db	2690	TCAGAGGCTGGAAGAAACAGTCTCCGGTGTACATAGGAAGAACAAACATAGAGACTTAATA	2749
Qy	399	-----	399
Db	2750	CATCGCGCGCTGGGATTCACGCTCCGGGATTCAGGGAGCATAGAGCAGCTCCCT	2809
Qy	400	---GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGln---	417
Db	2810	CTCTCTGCTCGGAAATTGACCCGACGCTTCTAGAAAAAGCCAAAAAG	2869
Qy	418	---SerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIle---IleAlaAlaPro	435

[illegible]

Qy	595	GlyLeuArgValHisPheAsnLeu-----ProGluAspAspLysGlySerGluAla	611
Db	4010	TCTGCTCCAAATCTGTATCCCTGGGAAATCTCTGTAGACAATAGTGT-----	4060
Qy	612	SerSerGluAlaGlyValValThrArgGlnThrArgProAspSer***GlnGluArg	631
Db	4061	-----CACCTGAGCATGCTGGGACCATCCAACCTGGGAAGATTCACTGGAA-----	4108
Qy	632	PheAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProPro---	650
Db	4109	-----ACAACACCATCTCCAGC-----CCCTCAGCACACCCCTCAATA	4147
Qy	651	ProValLeuAlaLysProLysLeuaspSerThrGlnLeuGlnGlnLeu-----HisAsn	668
Db	4148	CCACAGACCAAAATCTCAAGAGGAANAATCCCTGTCACCATGATCTTGTAAATAAC	4207
Qy	669	Gln-----ValLeuLeuGluGlnHisGlnLeu-----GlnAsnProPro	681
Db	4208	CAGAAGAAGGAGGGATGTTAAAGAATCCATATCAATTGGTGTACAAAGAACCAGCC	4267
Qy	682	ProSerSerProLys-----GluPheProPhe***Met	692
Db	4268	GCAAAGCTTCCCAAAATAGTCTCTTTTACCCACAGGTCAGAGTCCCCCTCAGATTCT	4327
Qy	693	ThrValLeuAsnSerAsnAlaProProAlaValThrThrSer-----	706
Db	4328	ACAACTCTCTGTACAGTCCGCCACACAGCTCTGTCTACACAANTGGTGCACCTCAGAAC	4387
Qy	707	-----***LysGlnValLysAlaProSerSerGlnThrPhe	718
Db	4388	AAGGGCACTGAAGTAGTATGATGAGTGGCCAGCAAGTCTCTCAGCAGGGAAGAAGCCCTTC	4447
Qy	719	SerLeuAlaArgProLysTyrPhePheProSerThr-----AsnThrThr	733
Db	4448	ACCAACTCTCTCTCA-----GTGCTTCTAGCACCATTAAGCAAGAGATCTAATACATTA	4501
Qy	734	-----AlaAlaThrValAlaProSer-----SerSerProValPheThrLeuSerSerThr	750
Db	4502	AACTTCTTGTCAACGGAAACCCCAACAGTGACAAGTCTCTACTGCTACTGCTCTGTCAAT	4561
Qy	751	ProGlnThrIleGlnArgThrValSerLysGluSer-----	762
Db	4562	ATGTCTGAACCCACCAAGCAAGATCCAAAGAAGAAAGAAAGACCAATAAAGGGCCCTCGG	4621
Qy	762	-----	762
Db	4622	AAGAACAGAAACACGCAACACACCCCCAGCAGGTTTCTGGCTATAGTGCATACTCA	4681
Qy	763	-----LeuLeuValSerHisProSerValGln-----	771
Db	4682	GCTCTAACACAGCTGATACCCCTTGGCTTTTCAGTCAATCCCCACGACAAGATGATGGT	4741
Qy	771	-----	771
Db	4742	GGAATGTAGTCAGTGTCTTATCACTCAACAACCTCTCTCTGGCCATCAACTGAAC	4801
Qy	772	-----Thr	772
Db	4802	TTTGAGAAGTACACCCAGACTTTGGGAAATACAAACAGCTTTTGGAAACAACGCTTGTGAGC	4861
Qy	773	LysSerProGlyGlyLeuSerIleGlnAsn-----GluProLeu-----	785
Db	4862	AAATCACAGGAGAGTACCACAGTGAAGAGAGGCTCAGACACACACCACTCCTCTCAGC	4921
Qy	786	-----ProProGlyProThrGluProThrProProPheThrPheSerIlePro	802
Db	4922	AGTGGGGGCCCCAGTGGCCACT---CCTTCCCAACCTCTTTTACTAAGGTGTGGTT	4978
Qy	803	SerGlyAsnGln-----PheGlnProArgCysValSerProIleProValSer	818
Db	4979	ACAGACAGCAAGATCAGATCAGCTTTCCAGATGACGCTCAAATAGAGTGTCCCATATAT	5038

```
QY 819 ProThrSerArg-----IleGlnAsnProValAla----- 828
Db : : : : :
QY 5039 GAATCTTCAAGGCACAATACAGATCTGCAGCAACCTTCAGCAGAGGCTAGGCCCAATCCT 5098
QY 829 -----PheLeuSerSerValLeuPro 835
Db : : : : :
QY 5099 GAGATCAATACTGGAAACCACTGACTCTCCCTCTAATCTGTTCATCCACCTCTGTGTGCCA 5158
QY 836 SerLeuProAlaIleProProThrAsnAlaMet**LeuProArgSerAlaProSerMet 855
Db : : : : :
QY 5159 GCACCTAAGGGTAGATAAACCCACAGAAATCTAAATGGAAGCCC----- 5200
QY 836 ProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAsn 875
Db : : : : :
QY 5201 -----TCTCCCTGGCCA----- 5212
QY 876 IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAsp 895
Db : : : : :
QY 896 ValArgProAsnGlnGlnGluThrLysIleSerSerPheGluGlnArgLeuMetAsnGlu 915
Db : : : : :
QY 5213 -----GACACAATATCAGCTCAAGTCATCACTCCGAACCACTTGAGAGGGC 5260
QY 916 IleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAsp 935
Db : : : : :
QY 5261 AAAGGCCAGCAGTAAGCATGTCCCCC-----CACCTC 5293
QY 936 GluIleProThrGlyLysCysIleAla----- 944
Db : : : : :
QY 5294 AGCCTTCAGAGGCCAGCAGCTCATGCCTCACACTGGAATACACAGACATGCAGAAAG 5353
QY 945 ProIlePheAspLysArg----- 950
Db : : : : :
QY 5354 AGTGTTTGTATAGAAACCTGGTCAAAACCCAACTTCCAAACACTCTCCCTTACGTCTCT 5413
QY 951 -----LeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
Db : : : : :
QY 5414 CTACCTAAGACTCTATTGAAAGCCAAAGAAATAATTGGAGAAAGGCTGCAAGCTTTACA 5473
QY 966 -----CysLysIleValGlyIleProValProLys 975
Db : : : : :
QY 5474 GTTCAGCTAATTCAGACGTTTCTTCTGTGTGAGGCTGTGGAGACCCACTGCCCATC 5533
QY 976 ValTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
Db : : : : :
QY 5534 ATCCACTGGACAGAGTTTCATCAGGANTTGAATATCCCAAGGCACACAGAAAGCCGG 5593
QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp 1013
Db : : : : :
QY 5594 TTCCACGTGCTTCCCAATGGCACC-----TTGTCCATCCAGAGGGTCAGTATTACAGGAC 5647
QY 1014 AspGlyAsnThrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis 1033
Db : : : : :
QY 5648 CGTGACAGTACTGTGCTGCTGCTGCTTAAATCCACTGGCGGTAGACCAATTTTCATGTCCT 5707
QY 1034 LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyLysSerHisArg 1053
Db : : : : :
QY 5708 TTGCTGTGGTTTTTACCG----- 5728
QY 1054 GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPheArgPro 1073
Db : : : : :
QY 5729 -----GCAAGGATTTTGGACAGACATGTCAAG----- 5755
QY 1074 HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuLysArgLeuAsp 1093
Db : : : : :
QY 5756 -----GAGATCACAGTTTCATTTTGGAAAGTACTGTGGAACATAAG 5794
QY 1094 CysLysValSerGlyLeuProProGluLeuThrTrpLeuAsnGlyGlnProVal 1113
Db : : : : :
QY 5795 TGCAGGTGGAGGTATCCCGAGGCTACGGTTCTCTGGATCTTGCACAAACCAACCGTG 5854
QY 1114 LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer 1130
```

```
Db 5855 GTCTCAGAAACGGCCAGGAAGCAGAAAGCTCTGGGTAAACACCTGATGGA-----ACA 5908
QY 1131 LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThrAsn 1150
Db : : : : :
QY 5909 TTGATCATCTAATCTAGCTTTTATGCTGCTTTTACAAAGTGTGTGGCCAGCAAC 5968
QY 1151 LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys 1170
Db : : : : :
QY 5969 CCATCTGCCAGGATTCATCTGTTGTTAAGATACAAGTCATCACA-----GCT 6016
QY 1171 AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro 1188
Db : : : : :
QY 6017 CCCCTGTGCTATTATAGACAAAGAGGCAAGCCATCGTTGGGTT---TTAGGTGGAAGT 6073
QY 1189 ValArgLeuGluCysArgValIleGlyMetProProValPheTrpTrpLysAsp 1208
Db : : : : :
QY 6074 TTGAACCTGCCCTGCACCTGCAAAAGAACTCCCGAGCTAGTGTTCACCTGGGTCCTTTAT 6133
QY 1209 AsnGluThr-----IleProCysThrArgGluArgIleSerMetHisGlnAsp 1224
Db : : : : :
QY 6134 GATGGACTGAACATAAACCACTTCAGTTGACTCATTCACAGATTTTCTGTATCCAAAT 6193
QY 1225 ThrThrGlyTyAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTrp 1244
Db : : : : :
QY 6194 GGAAC-----CTGTATATAAGAACATCGCTCTTCAGTGAGGGCAGCTTAT 6241
QY 1245 ThrLeuSerAlaLysAsnGluAlaGly----- 1253
Db : : : : :
QY 6242 GAGTGCAATTCACACAGCTCCTCAGGCTCAGAGAGAGGGTAGTGATTTCTTACTGTGGA 6301
QY 1253 ----- 1253
Db 6302 GAGGAGAGACAATCCCGAGGATAGAAACTGCCTCTCAGAATGGAGCTGAGTGAATTTG 6361
QY 1254 -----IleValSerCysThrAlaArgLeuAsp----- 1262
Db : : : : :
QY 6362 GGTGAGAAATTAATACTGCACTGCTCAGCTACTGGGATCCAAAGCCCTAGAAATAATCTGG 6421
QY 1263 -----IleTyAlaGlnTrpHisGlnIleProProMetSer 1276
Db : : : : :
QY 6422 AGGTGTCATCCAGGCTGTCATCGACCACTGAGGAGAGAGGGTAGTGATTTCTTACTGTG 6478
QY 1277 ValArgProSerGlySerArg---TyrGlySerLeuThrSerLys 1290
Db : : : : :
QY 6479 GTCTACCAAAATGATCCTTGTGTTGGGTGAGTACGAGGAGAAA 6523
RESULT 14
US-09-991-630-1
; Sequence 1, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS TH
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US/09/991,630
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8883
; TYPE: DNA
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(8883)
```

; OTHER INFORMATION: 'n' can be any nucleotide 'a', 'c', 'g' or 't'.
US-09-991-630-1

Alignment Scores:		6.73e-19	Length:	8883
Pred. No.:	416.50	Matches:	354	
Percent Similarity:	30.73%	Conservative:	216	
Best Local Similarity:	19.08%	Mismatches:	552	
Query Match:	6.04%	Indels:	733	
DB:	10	Gaps:	83	
US-09-818-990B-2 (1-1320) x US-09-991-630-1 (1-8883)				
QY	2	GlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu	18	
DB	1460	CAGGAGCAATGGATCGCTCCACCTCACCTCAAGATTTCATAGAACCTTTGGTCC	1519	
QY	19	SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArgAla	35	
DB	1520	TTGTCTTTTGAACATGACANANNTNTCTGGAATAAGCCGACATGGTCTGTAGTATCCAA	1579	
QY	36	GluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly	55	
DB	1580	AAGCCATCAAGGACA-----TCACCAACTGCATTCACTGAA-----	1615	
QY	56	GlyGlnAspAspLeuProAspLeuSerAlaPheLeuSerGlnGluGluLeuAspGluSer	75	
DB	1616	---GAAATGACTACATCATGCTTAATGCGTCATT-----TCC	1651	
QY	76	ValAsnLeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGln	95	
DB	1652	ACAAATCTTGTG---TGCAGTGTAGATTATAATCACATCCAGCAGGTGGCAACTCTG	1708	
QY	96	AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluPro	115	
DB	1709	GCTTTATACAGTGACTCTCTCTGATAGTAAAGAGGCCGCCAGCTTACC-----	1759	
QY	116	AsnPheCysGlnAspAsnProArgSerProThrSerLysGluSerProGlnGluAla	135	
DB	1760	-----GAGACTCCTTCACTGTCTTCTAGATATAACAGGTGCT	1798	
QY	136	LysArgProGlnTyrCysSerGluThr-----	144	
DB	1799	CTTAGGCTGAAGACATTTTACCAGCATAGAGGCTGATGTCAGAGCAGACCCCTTTTGG	1858	
QY	145	---GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluLeu	161	
DB	1859	TTCCAACAAGAAAAATTGCTTCGAGCTGAACAGAACTGCCACCACACATTAGCACATTA	1918	
QY	162	SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis	181	
DB	1919	CAGATCCAGTTTTCACATGATGCTCAATCGCTTTA---CCAAGGCG---GAGATGAGA	1972	
QY	182	LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSerAsp	201	
DB	1973	CGCGAGAGACTCAATGGACCATGATCCTGTATGATGAACAAT-----	2014	
QY	202	LeuSerGluArgArgGluArgSer-----SerValPro	212	
DB	2015	-----CCCAAACTGGAACGACGTCTCTGTTGGCGGCACATTATGCCCTGAGCTGCCA	2068	
QY	213	-----IlePro	214	
DB	2069	GGCAAGGCGGCCCTTCCACCTCAGTGGATGGCTTCTAGCTGATGGAGTAAGTGAGA	2128	
QY	215	IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln	230	
DB	2129	GCCCTTACGTTAGCGAGGTGGCGGAATCCTAATAGACAAAAATGGGAAGTTGGAAGT	2188	
QY	231	GlnGluAlaLysArgArgGluAla-----GluGlnAlaAlaSerGluAlaAlaGly	247	
DB	2189	CAGATGGCTGACAGCTTTGATGTCAGGCTTTTACCACCTGCATGAACCAATGATGCAGAT	2248	

QY	248	GlyAspThrThrProGlySerSerProSerSerLeuTyrTyrGlu-----Glu	263	
DB	2249	CGCGATGTT-----CTCACATACAGGATAACTGGTAGAG	2284	
QY	264	ProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGly	283	
DB	2285	CCCTATGGAGAAAGCACACATGACAGTGGAGTC-----CAGCACACAGTGGTACGGGT	2338	
QY	284	ThrArgValGlnLeuAspCysIleValIleValIleProProGlnValArgTyrTyr	303	
DB	2339	GAGACGCTCGACCTTCATGCCCTTCCAGGGTGTCCAGATGCTTCTATTAGCTGAT	2398	
QY	304	CysGluGlyLysGluLeuAsnSerPro-----AspIleHisIleValGlnAlaGly	321	
DB	2399	CTTCCAGGGAACACTGTTCTCTCAGCCATCAAGAGACAGGCAAAATCTTAAACAATGG	2458	
QY	322	AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys	341	
DB	2459	ACCTTAAGAATATTACAGGTTAGC-----CCAAAGATCAAGTCTATTACCAATGT	2509	
QY	342	PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIle-----	358	
DB	2510	GTGCTGCCAACCCATCAGGGCGGCTTTCCAGTTTAAAGTTTCAGTTCAAAAGAAA	2569	
QY	359	---GluGlyValSerSerAspSerGlu-----GlyAspProAsn	371	
DB	2570	GGCCAAAGGATGTTGAGCATGACAGGGAGGAGGTGGATCGACTGGAGAACCCAC	2629	
QY	372	LysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerProThrThrSer	391	
DB	2630	TCCAGTGTTCCTTAAAGCAGCCAGCATCTTTGAACATCTCTGCATCAGCTTTGACAGG	2689	
QY	392	AlaValIleProProAlaValPro-----	399	
DB	2690	TCAGAGCTGGAACAAAGTCTCCGGTGTACATAGGAAGAACAAACATAGAGCTTAATA	2749	
QY	399	-----	399	
DB	2750	CATCGGGCGGTGGGATTCACCGCTCCGGGATTCAGGGAGCATAGGAGGCAGCTCCCT	2809	
QY	400	---GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGln---	417	
DB	2810	CTCTCTCTCGGAGATTGACCCGCAACGCTGGCGAGCACTTCTAGAAAAAGCCAAAAAG	2869	
QY	418	---SerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIle---IleAlaAlaPro	435	
DB	2870	ANTTCTGTGCCAAAAAGCAAGAAATACCACAGTAAGCCAGTGCCTGCTGCTTCC	2929	
QY	436	ValPheThrLysMetLeuGlnAsnLeuSerAlaSer-----GluGlyGln	450	
DB	2930	CTCGTGGAACTCACTGACGAGGAAAGAGATCCCTCTGCGATGATTCCTCCAGATGAAGAA	2989	
QY	451	LeuValValPheGluCysArgValLysGlyAlaPro-----SerProLysValGlu---	467	
DB	2990	TTTATGTTTCTGAAACAACTAAGGCTTCTGCTCCAGGAAGTCCCAACTGCTGACTCT	3049	
QY	468	-----TrpTyrArgGluGlyThrLeuIle-----	475	
DB	3050	GGACCACTAAATCATGTTTATGACGAGTATAGCTTCTGTCGACAGAGTCTCAACTGTG	3109	
QY	476	-----GluAspSerProAspPheArg-----	482	
DB	3110	AATCCACAACACTACAAATCTGACACCTTCTCTGATTTCAAAATTTTAGTGTAAACAAC	3169	
QY	483	-----IleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThr	499	
DB	3170	GGTACAGCTGTGACAAAAGAGATGAACCCATCCATAGCAAGCAAAATAGAAAGATACAAC	3229	
QY	500	-----LeuValIleAlaGluValPheAlaGlu-----AspSer-----	510	
DB	3230	AACCAAAACCAATCATTATCTTCCATCAGTAGTGAATTCGAGATTTCTGCTCAGCA	3289	
QY	510	-----	510	

Db	3290	GGAAGACCATCTTCCCAAAGTGCACACACCCTGTAAACAGGGGGAAACATGGCTACCTATGGC	3349
Qy	511	-----GlyCysPheThrCysThrAlaSer- 	518
Db	3350	CATACCAACACATATAGTAGCTTACCACGAAGCCAGTACAGTCTTGCACCAATAAAT	3409
Qy	519	-----AsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGlu	536
Db	3410	CCAACAGAAAGTTATGGACCTCAGATACCTATTACAGAGGTACAGCAGACCTAGCAGTAGT	3469
Qy	537	AspLeuSerAsnAsnGly-----SerLeuHisSerAlaAsnSerThr	550
Db	3470	GACATCTCTTCCACACTACTCGAGACCCCTAGCTTCTCCAGTCACCCCTTCAGGTTCAAC	3529
Qy	551	ThrAsnLeuAlaIleGluProGlnPro-----	560
Db	3530	ACCATGCCCTCGCTTTATTTCACATTCCTAGAACACAATACAGGTAACTTCCCTTGG	3589
Qy	561	-----SerProProHis	564
Db	3590	TCCAGGCACTTGGGAAGAGAGAGGACAAATTTGGAGCAGAGGGAGAGTTAAAAACCCACAT	3649
Qy	565	SerGluPro-----	567
Db	3650	AGAACCCCACTTCCGAGCGGATACACAGGACTGTGAGCCAGCAATCAAGGGACCT	3709
Qy	568	-----ProSerValGluGlnPro-----	573
Db	3710	GCTAACCAAAATGTGAGCCCAAGTTCAGCCACAGAGTACCCTGGGATGTGCCACACATGT	3769
Qy	573	-----	573
Db	3770	CCTTCGCAGAGGGGCTCACAGTGGCTACTGCAGCACTGTTCAGTTCCAAGTTATCCAC	3829
Qy	574	-----ProlysProlysLeuGluGlyValLeuVal-----	583
Db	3830	AGTGCCTCCCCAAAATAATATGTTGGGTTCATAGCAGAAAGAGTCTACCACTGTGTC	3889
Qy	583	-----	583
Db	3890	AGNAACCACTGTACTATTTAAGCAACAACAAAATGTAGATATTTGAGATAAATCAACAACC	3949
Qy	584	-----AsnHisAsnGluProArgSerSerSerArgIle	594
Db	3950	ACTACAAAATATTCGGAGGGGAAAGTAACCAAGTGTCTCTACGGAAGCAAGCATGACT	4009
Qy	595	GlyLeuArgValHisPheAsnLeu-----ProGluAspAspLysGlySerGluAla	611
Db	4010	TCTGCTCCAACATCTGTATCCCTGGGAAATCTCTGTAGACAATAGTGT-----	4060
Qy	612	SerSerGluAlaGlyValValThrThrArgGlnThrArgProAspSer***GlnGluArg	631
Db	4061	-----CACCTGAGCATGCTGGGACCATCCAAACTGGGAAGATTCAGTGGAA-----	4108
Qy	632	PheAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProPro-----	650
Db	4109	-----ACAACCACTTCCCAGC-----CCCTCAGCAGCACCCCTCAATA	4147
Qy	651	ProValLeuAlaLysProlysLeuAspSerThrGlnLeuGlnGlnLeu-----HisAsn	668
Db	4148	CCAACAGACACAAAATTCFAAGAGGAGAAAACCTCCCTTGCACAGATCTTGTAAATAAC	4207
Qy	669	Gln-----ValLeuLeuGluGlnHisGlnLeu-----GlnAsnProPro	681
Db	4208	CAGAAGAGGAGGGGTGTTAAAGAAATCCATATCAATTCGGTTTACAAAAGAACCCAGCC	4267
Qy	682	ProSerSerProlys-----GluPheProPhe***Met	692
Db	4268	GCAAAGCTTCCCAAAATAGCTCCTCTTTTACCACAGGTCAGAGTTCCCCCTCAGATTCT	4327
Qy	693	ThrValLeuAsnSerAsnAlaProProValValThrThrSer-----	706

[illegible]

```
QY 936 GluileProThrGlyLysCysIleAla----- 944
:::|::|
Db 5294 AGCCTCCAGAGCCAGCACTCATGCTCACACTGGAATACACAGAGCATGCAGAAAAG 5353
QY 945 ProIlePheAspLysArg----- 950
:::|::|::|::|::|
Db 5354 AGTGTTTTTGATAAGAAAGCTGGTCAAAACCCCAACTTCCAAACATCTGCCTTACGCTCTCT 5413
QY 951 -----LeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
|::|::|::|::|::|::|
Db 5414 CTACCTAAGACTCTATTGAAAGACCAAGAAATAATTGGAGGAAGGCTGCAAGCTTTACA 5473
QY 966 -----CysLysIleValGlyIleProValProLys 975
|::|::|::|::|::|::|
Db 5474 GTTCCAGCTAATTCAGACGTTTTTCTCTCTGTGAGGCTGTGGACACCCACTGCCCATC 5533
QY 976 ValTyrTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
:::|::|::|::|::|::|::|
Db 5534 ATCCACTGGACCAAGAGTTTCATCAGGANTTGAATATCCCAAGGGACACAGAAAAGCCGG 5593
QY 994 MetArgArgGluClyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp 1013
:::|::|::|::|::|::|::|
Db 5594 TTCCACGTGCTTCCCAATGCGACC-----TTGTCCATCCAGAGGTCAGTATTCAGGAC 5647
QY 1014 AspClyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis 1033
|::|::|::|::|::|::|::|
Db 5648 CGTGGACAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5707
QY 1034 LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArg 1053
|::|::|::|::|::|::|::|
Db 5708 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5728
QY 1054 GlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgPro 1073
:::|::|::|::|::|::|::|
Db 5729 -----GCAAGGATTTTGGACACATGTCAG----- 5755
QY 1074 HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuLeuAsp 1093
|::|::|::|::|::|::|::|
Db 5756 -----GAGATCACAGTTTCACTTGGAGTACTGTGGAACTAAAG 5794
QY 1094 CysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVal 1113
|::|::|::|::|::|::|::|
Db 5795 TGCAGAGTGAGGGTATGCGAGCGCTACGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5854
QY 1114 LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer 1130
:::|::|::|::|::|::|::|
Db 5855 GTCTCAGAACGCCCAAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5908
QY 1131 LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsn 1150
|::|::|::|::|::|::|::|
Db 5909 TTGATCATCTAATCTGAGCTCTTATGATCGTGGTTTACAGTGTTGGCCACCAAC 5968
QY 1151 LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLys 1170
:::|::|::|::|::|::|::|
Db 5969 CCATCTGGCCAGGATTCACCTGTTGGTTAAGATAACAAGTCATACA-----GCT 6016
QY 1171 AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro 1188
|::|::|::|::|::|::|::|
Db 6017 CCCCTGCTCATATAGACAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6073
QY 1189 ValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAsp 1208
:::|::|::|::|::|::|::|
Db 6074 TTGAACCTGCCCTGCACTGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6133
QY 1209 AsnGluThr-----IleProCysThrArgGluArgIleSerMetHisGlnAsp 1224
:::|::|::|::|::|::|::|
Db 6134 GATGGGAGTGAACATAAACCACTTGCAGTTTCCAGATTTTCTTGATATCCAAAT 6193
QY 1225 ThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTyrTyr 1244
|::|::|::|::|::|::|::|
Db 6194 GGAAC-----CTGTATATAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6241
```

```
QY 1245 ThrLeuSerAlaLysAsnGluAlaGly----- 1253
|::|::|::|::|::|::|
Db 6242 GAGTGCATTGCCACCAGCTCCTCAGGCTCAGAGAGAGGTAGTGATTCTTACTGTGGAA 6301
QY 1253 ----- 1253
Db 6302 GAGGAGAGACAATCCAGGATAGAAACTGCCTCTCAGAAATGGACTGAGGTGAATTGG 6361
QY 1254 -----IleValSerCysThrAlaArgLeuAsp----- 1262
:::|::|::|::|::|::|::|
Db 6362 GGTGAGAAATTACTACTGAAGTCTCAGCTACTGGGATCCAAAGCCTAGAATAATCTGG 6421
QY 1263 -----IleTyrAlaGlnTrpHisHisGlnIleProProPrometSer 1276
|::|::|::|::|::|::|::|
Db 6422 AGGCTGCCATCCAAAGCTGTCATCGACAGTGG--CACAGATGGCGACGCCGAATCCAC 6478
QY 1277 ValArgProSerGlySerArg---TyrGlySerLeuThrSerLys 1290
|::|::|::|::|::|::|::|
Db 6479 GTCTACCCAAATGGATCTTGGTGGTGGTCTAGTGACGGAATAA 6523
```

RESULT 15

```
US-09-991-630-5
; Sequence 5, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS TH
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT APPLICATION NUMBER: US/09/991,630
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patent version 3.0
; SEQ ID NO 5
; LENGTH: 8883
; TYPE: DNA
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(8916)
; OTHER INFORMATION: n can be any amino acid
US-09-991-630-5
```

Alignment Scores:

```
Pred. No.: 6,73e-19 Length: 8883
Score: 416.50 Matches: 354
Percent Similarity: 30.73% Conservative: 216
Best Local Similarity: 19.08% Mismatches: 552
Query Match: 6.04% Indels: 733
DB: 10 Gaps: 83
```

US-09-818-990b-2 (1-1320) x US-09-991-630-5 (1-8883)

```
QY 2 GlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArg-----Glu 18
|::|::|::|::|::|::|::|
Db 1460 CAGGAGACAATGATGCTGCCTCCACTCACCAAGATTTCATAGAACCTTTGGCTCC 1519
QY 19 SerTyrLeuAlaGluThrArgHisArgGlyAsnAsnGluArg-----SerArgAla 35
|::|::|::|::|::|::|::|
Db 1520 TTGCTCTTGAACATGACANANNTTCTGGAATAAAGCCGACATGGTCTGTAGTATCCAA 1579
QY 36 GluProSerSerAsnProCysHisPheGlySerProSerGlyAlaAlaGluGlyGly 55
:::|::|::|::|::|::|::|
Db 1580 AACCCATCAAGGACA-----TCACCAACTGCATTCAGTAA----- 1615
QY 56 GlyGlnAspLeuProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSer 75
:::|::|::|::|::|::|::|
```

Db	1616	---GAAANTGACTACATCATGCTAAATCGCGTCATTT-----TCC	1651
Qy	76	ValAsnLeuAlaArgLeuAlaIleAsnTyArgProLeuGluLysAlaAspGluThrGln	95
Db	1652	ACAAATCTTG---TCCAGTGTAGATTATATCATCCAGCCAGCTGGCACTCTCTG	1708
Qy	96	AlaArgLysArgLeuSerProAspGlnMetLysHisSerProAsnLeuSerPheGluPro	115
Db	1709	GCTTTATACAGTGACTCTCCTCTGCTACTAGAAAGGAAGCCCGACTTACC-----	1759
Qy	116	AsnPheCysGlnAspAsnProArgSerProThrSerLysGluSerProGlnGluAla	135
Db	1760	-----GAGACTCCTTCACCTGCTCTTCTAGATATAACAGAGTGCGCT	1798
Qy	136	LysArgProGlnTyrcysSerGluThr-----	144
Db	1799	CTTAGCGCTGAAGACATTTTACCAGCATAGAGGCTGATCTCAGAGCAGACCCCTTTTGG	1858
Qy	145	---GlnSerLysLysValPhe-----LeuAsnLysAlaAlaAspPheIleGluGluLeu	161
Db	1859	TTCCAACAAGAAAAAATTCTTGCAGCTGAACAGAACTGCCACCACACTTAGCACATTA	1918
Qy	162	SerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsnHis	181
Db	1919	CAGATCCAGTTTTTCCACTGATGCTCAAAATCGCTTTA---CCAAGGGG---GAGATGAGA	1972
Qy	182	LysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSerAsp	201
Db	1973	GCGGAGAGACTCAAAATGGACCATGATCTGTATGATGAACAAT-----	2014
Qy	202	LeuSerGluArgArgGluArgSer-----	212
Db	2015	---CCCAAACTGGAACGCACCTGCTCGTGGTGGCGGCACATAITGCCCTGAGCTGTCCA	2068
Qy	213	-----IlePro	214
Db	2069	GGCAAAAGGGGACCCCTTACCTCACTTGGAAATGGCTTCTAGCTGATGGAGTAAAGTGAGA	2128
Qy	215	IleProAlaAspThrArgAspAsnGluVal-----AsnHisAlaLeuGluGln	230
Db	2129	GCCCCCTTAGTCGAGGATGGGCGAATCCTAATAGACAAAAATGGAGTTGGAACTG	2188
Qy	231	GlnGluAlaLysArgArgGluAla-----GluGlnAlaAlaSerGluAlaAlaGly	247
Db	2189	CAGATGGCTGACAGCTTTCATGCAGCTCTTACCACCTGCATAAGCAACCATGATGCAGAT	2248
Qy	248	GlyAspThrThrProGlySerSerProSerSerLeuTyTyrglu-----Glu	263
Db	2249	GCGGATGTT-----CTCACATACAGGATAAAGTGTGGTAGAG	2284
Qy	264	ProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGly	283
Db	2285	CCTATGGAGAAGAACACATGACAGTGGAGTC-----CAGCACACAGTGGTTACGGGT	2338
Qy	284	ThrArgValGlnLeuAspCysIleValValGlyIleProProGlnValArgTyrTyr	303
Db	2339	GAGACGCTCGACCTTCCATCGCTTTCCACGGGTGTTCAGATGCTTCTATTAGCTGGATT	2398
Qy	304	CysGluGlyLysGluLeuGluAsnSerPro-----AspIleHisIleValGlnAlaGly	321
Db	2399	CTTCCAGGGAACACTGTGTCTCTCAGCCATCAAGACAGACAGGCAAAATTTTAAACAATGGG	2458
Qy	322	AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrcys	341
Db	2459	ACCTTAGAATATTACAGGTTACG-----CCAAAGATCAAGGTCAATTACCAATGT	2509
Qy	342	PheAlaSerAsnIleTyrglyThrAspSerThrSerAlaGluIleTyrlle-----	358
Db	2510	GTGGCTGCCAACCCATCAGGGGCCACACTTTTCCAGTTTAAAGTTTCAGTTTCAAAAGAAA	2569
Qy	359	---GluGlyValSerSerAspSerGlu-----GlyAspProAsn	371
Db	2570	GGCCAAGAGTGTGTGAGCATGACAGGAGCGCAGGTGGATCTGGACTTGGAGAACCAAC	2629

Qy	372	LysGluGluMetAsnArgIleGlnLysProAsnGluValSerSerProProThrThrSer	391
Db	2630	TCCAGTGTTCCTTAAAGCAGCCAGCATCTTTGAAACTCTCTGCATCAGCTTGACAGG	2689
Qy	392	AlaValIleProProAlaValPro	399
Db	2630	TCAGAGGCTGGAAACAAGTCTCCGGTGATACATAGGAAGAACAACATAGAGACTTAATA	2749
Qy	399		399
Db	2750	CATCGCGCGCTGGGATTCCACGCTCCGGCGATTTCAGGAGCATAGGAGGAGCTCCCT	2809
Qy	400	---GlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGln---	417
Db	2810	CTCTCTGCTCGGAAATTGACCGCAACGCTGGCGCAGCATTTCTAGAAAAAGCCAAAAG	2869
Qy	418	--SerProThrAsnTyrLeuGlnGlyLeuAspGlyLeuProIle---IleAlaAlaPro	435
Db	2870	AATTCTGTGCCAAAAAGCAAGAAATACCACAGTAAGCCAGTGCCTAGCTGGCTGTTCC	2929
Qy	436	ValPheThrLysMetLeuGlnAsnLeuSerAlaSer---	450
Db	2930	CTCTGGAACTCACTGACGAGGAAAAGGATGCTCTGGCATGATTCCTCCAGATGAAGA	2989
Qy	451	LeuValValPheGluCysArgValLysGlyAlaPro-----SerProLysValGlu---	467
Db	2990	TTTCATGTTCTGAAAACTAAGGCTTCTGGTGCCCGAAGGTCACCAACTGCTCACTCT	3049
Qy	468	-----TrpTyrArgGluGlyThrLeuIle-----	475
Db	3050	GGACCATTAATCATGTTTATGACGAGTAGTCTTGCCACAGAGTCTCAACTGTG	3109
Qy	476	-----GluAspSerProAspPheArg-----	482
Db	3110	AATCCACAACACTACAACTCTGAGCAGCTTCCTGATTTCAAATTTATTTAGTGAACAAC	3169
Qy	483	-----IleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThr	499
Db	3170	GGTACAGCTGTGACAAAGATGATGACCCATCCATAGCAGCAAGCAAAATAGAGATACAAC	3229
Qy	500	-----LeuValIleAlaGluValPheAlaGlu-----AspSer-----	510
Db	3230	AACCAAAACCAATCATATTATCTTCCATCAGTAGTGAAATTCGAGATTCTGCTCAGGCA	3289
Qy	510		510
Db	3290	GGAAGAGATCTCCCAAGTCGACACCCTGTACAGGGGGAAACATGGCTACCTATGGC	3349
Qy	511	-----GlyCysPheThrCysThrAlaSer-----	518
Db	3350	CATACCAACACATATAGTACGTTTACCAGCAAGCCAGTACAGTCTTCCAGCCCAATAAT	3409
Qy	519	-----AsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGlu	536
Db	3410	CCAAACAGAAAGTTATGGACCTCAGATACCTATTATACAGGAGTCAGCAGACCTAGCAGTAGT	3469
Qy	537	AspLeuSerAsnAsnGly-----SerLeuHisSerAlaAsnSerThr	550
Db	3470	GACATCTCTCTCACTACTGACGACCCCTAGCTTCTCCAGTCACCCCTCAGGTTTCACAC	3529
Qy	551	ThrAsnLeuAlaIleGluProGlnPro-----	560
Db	3530	ACCAGCTCGCTTATTTCACATTCCTAGAAACAACATACAGGTAACTTCCCTTGG	3589
Qy	561	-----SerProProHis	564
Db	3590	TCCAGGCACTTGGAGAGAGAGGACAATTTGGAGCAGAGGGAGAGTTAAAAACCCACAT	3649
Qy	565	SerGluPro-----	567
Db	3650	AGAACCCCACTTCCGAGCGCATAGACACAGGACTGTGAGGCCAGCAATCAAGGGACCT	3709

QY 568 -----ProSerValGluGlnPro----- 573
Db 3710 GCTAACAAAAATGTGAGCCAAGTTCCAGCCACAGAGTACCCTGGGATGTGCCACACATGT 3769
QY 573 ----- 573
Db 3770 CCTTCCGACAGAGGGCTCACAGTGGCTACTGTCAGCACTGTCTCAAGTTTCATCCAC 3829
QY 574 -----ProLysProLysLeuGluGlyVal----- 583
Db 3830 AGTGCCCTCCCAAACTAATATGTGGGTCTATAGCAGAAAGTCTACCACTGTGGTC 3889
QY 583 ----- 583
Db 3890 AAGAAACCACTGTTACTATTAAAGCAACAATAAGTATAGATATTGAGATAATAACAACC 3949
QY 584 -----AsnHisAsnGluProArgSerSerSerArgIle 594
Db 3950 ACTACAAAATATTCGGAGGGGAAGTAACCGTGTATCTCTACGGAAGCAAGCATGACT 4009
QY 595 GlyLeuArgValHisPheAsnLeu-----ProGluAspAspLysGlySerGluAla 611
Db 4010 TCTGCTCCACATCTGTATCCCGAGGGGAATCTCTCTAGACAAATAGTGT----- 4060
QY 612 SerSerGluAlaGlyValValThrArgGlnThrArgProAspSer***GlnGluArg 631
Db 4061 -----CACCTGAGCATCCCTGGGACCATCCAACTGGGAAGATTCAGTGGAA----- 4108
QY 632 PheAsnGlyGlnAlaThrLysThrProGluProSerPheProValLysGluProPro--- 650
Db 4109 -----ACAACACCACTTCCAGC-----CCCTCAGCACACCTCAATA 4147
QY 651 ProValLeuAlaLysProLysLeuAspSerThrGlnLeuGlnLeu-----HisAsn 668
Db 4148 CCAACAGCACAAAATCTCAAGAGAGAAACTCCCTGTGCACCATGCTTTGTAAATTAAC 4207
QY 669 Gln-----ValLeuLeuGluGlnHisGlnLeu-----GlnAsnProPro 681
Db 4208 CAGAAGAAGAGGGGATGTTAAAGAATCCATATCAATTCGGTTTACAAAAGAACCCAGCC 4267
QY 682 ProSerSerProLys-----GluPheProPhe***Met 692
Db 4268 GCAAGCTTCCCAAAATFAGTCTCTTTTACCACAGGTGAGAGTTCCTCCCTCAGATTCT 4327
QY 693 ThrValLeuAsnSerAsnAlaProProAlaValThrThrSer----- 706
Db 4328 ACAACTCTTGACAAGTCGCGCCACAGCTGTGTCTACAAATGCTGCTGCTGCTGCTGCTGCT 4387
QY 707 -----***LysGlnValLysAlaProSerSerGlnThrPhe 718
Db 4388 AAGGGCACTGAAGTAGTATCAGGTGCGCAGAAGTCTCTCAGCAGGGAAGAACGCCCTTC 4447
QY 719 SerLeuAlaArgProLysTyrPhePheProSerThr-----AsnThrThr 733
Db 4448 ACCAATCTCTCTCA-----GTGCTTCTCAGCACCAATAAGCAAGATCTAATACATTA 4501
QY 734 -----AlaAlaThrValAlaProSer---SerSerProValPheThrLeuSerSerThr 750
Db 4502 AACTCTTGTCACAGGAAACCCCAACAGTACAGTCTACTGTCTACTGCTGCTGCTGCTGCT 4561
QY 751 ProGlnThrIleGlnArgThrValSerLysGluSer----- 762
Db 4562 ATGCTGAACCCCAACGAAGATCCAAAGAACCAAGACCAAAATAAAGGGCCCTCGG 4621
QY 762 ----- 762
Db 4622 AAGAACAGAAACACGCAAAACACCCAGCGAGGTTTCTGCTATAGTGCATACTCA 4681
QY 763 -----LeuLeuValSerHisProSerValGln----- 771
Db 4682 GCTCTAACACAGCTGTATACCCCTTGGCTTTCAGTCAATCTCCCCACGACAAGATGATGGT 4741
QY 771 ----- 771

Db 4742 GGAATGTAAAGTGCAGTGTCTTATCTACTCAACAACCTCTCTCTGGCCATAACTGAACGT 4801
QY 772 -----Thr 772
Db 4802 TTTGAGAAGTACACCCAGACTTTTGGGAAATACAAACAGCTTTGGAAACAACAGTGTGTGAGC 4861
QY 773 LysSerProGlyGlyLeuSerIleGlnAsn-----GluProLeu----- 785
Db 4862 AAATCACAGAGAGTACCACAGTGAAGAGCCCTCACACACACACACCTCTCTCAGC 4921
QY 786 -----ProProGlyProThrGluProThrProProPheThrPheSerIlePro 802
Db 4922 AGTGGGCGCCCGCCAGTGCCCACT---CCTTCCCACTCTCTTTACTAAGGTGTGGTT 4978
QY 803 SerGlyAsnGln-----PheGlnProArgCysValSerProIleProValSer 818
Db 4979 ACAGACAGCAAGTACATCAGCTTTCAGATGACGTCAATAGAGTGGTCACCATATAT 5038
QY 819 ProThrSerArg-----IleGlnAsnProValAla----- 828
Db 5039 GAATCTTCAAGGCACAATACAGATCTCGCAACCCCTCAGCAGAGGTAGCCCCAATCCT 5098
QY 829 -----PheLeuSerSerValLeuPro 835
Db 5099 GAGATCATAACTGGAACCACTGACTCTCCCTCTAATCTGTTCATCCACTCTCTGTGCCA 5158
QY 836 SerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMet 855
Db 5159 GCATAGGGGTAGATAAACCAACACAGAAATCTAAATGGAAGCCC----- 5200
QY 856 ProSerGlnGlyLeuAlaLysAsnThrLysSerProGlnProValAsnAspAsn 875
Db 5201 -----TCTCCCTGGCCA----- 5212
QY 876 IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysIleThrPheSerAsp 895
Db 5212 ----- 5212
QY 896 ValArgProAsnGlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGlu 915
Db 5213 -----GAACACAAATATCAGCTCAAGTCATACTCCGAAACCATTTGAGAAGGGC 5260
QY 916 IleGluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAsp 935
Db 5261 AAAAGGCCAGCATGTAACATGTCCTCC-----CACCTC 5293
QY 936 GluIleProThrGlyLysCysIleAla----- 944
Db 5294 AGCCTTCCAGAGGCCAGCACTCATGCTCCTCAGCTGGAATACACAGAAAGCATGCAAAAAG 5353
QY 945 ProIlePheAspLysArg----- 950
Db 5354 AGTGTGTTTGTATAAGAAACCTGGTCAAAACCCAACTTCCAAACATCTGCCCTAGCTCT 5413
QY 951 -----LeuLysHisPheArgValThrGluGlySerProValThrPheThr 965
Db 5414 CTACCTAAGACTCTATTGAAAACCCAGAAATAATTGGAGGAAGGCTGCAAGCTTTACA 5473
QY 966 -----CysLysIleValGlyIleProValProLys 975
Db 5474 GTTCCAGCTAATTCAGAGCTTTTCTCTTGTGGGTGTTGGAGACCCCACTGCCCATC 5533
QY 976 ValTyrTrpPheLys-----AspGlyLysGlnIleSerLysArgAsnGluHisCysLys 993
Db 5534 ATCCACTGGACACAGATTTTCATCAGANTTGAATATATCCCAAGGACACACAAAAGCGG 5593
QY 994 MetArgArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAsp 1013
Db 5594 TTCACAGTGTTCCTCAATGGCACCC-----TTGTCCATCCAGAGGGTCAGTATTTCAGGAC 5647
QY 1014 AspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlnArgIleSerCysSerGlyHis 1033

Db 5648 CGTGACAGTACCTGTGCTCTGCATTTAATCCACTGGCGGTAGACCAATTTTCATGTCTCT 5707
Qy 1034 LeuMetValGlnSerLeuProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArg 1053
Db 5708 TTGCTGTGGTTTTTTTACCG----- 5728
Qy 1054 GlyArgSerArgValGlnGlnArgAspLysGluProLeuGlnGluArgPheArgPro 1073
Db 5729 -----GCAAGGATTTTGACAGACATGTCAG----- 5755
Qy 1074 HisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAsp 1093
Db 5756 -----GATCACAGTTCATTTTGAAGTACTGTGGAACATAAG 5794
Qy 1094 CysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProVal 1113
Db 5795 TGCAGGTGGAGGTATGCCAGGCTACGGTTCTCGTACTTGCACAAACCAACGGTG 5854
Qy 1114 LeuProAsp-----AlaSerHisLysMetLeuValArgGluThrGlyValHisSer 1130
Db 5855 GTCTCAGAAACGGCCAGGGAAGCAGAAAGTCTGGGTAACACCTGATGGA-----ACA 5908
Qy 1131 LeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsn 1150
Db 5909 TTGATCATCTATAATCTGAGCTTTTATGATCGTGGTTTACAAAGTGTGGCCAGCAAC 5968
Qy 1151 LysThrGlyGlnAsnSerPheSerLeuGluLeuSerValAlaLysGluValLysLys 1170
Db 5969 CCATCTGCCAGGATTCATCTGTTGTTAAGATACAAGTCATCACA-----GCT 6016
Qy 1171 AlaProValIleLeuGluLysLeuGlnAsnCys-----GlyValProGluGlyHisPro 1188
Db 6017 CCCCTGTCTATTAGAGCAAAAGAGCGCAAGCCATCGTTGGGTT---TTAGTGGAAGT 6073
Qy 1189 ValArgLeuGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAsp 1208
Db 6074 TTGAAACTGCCCTGCCTCACTGCCAAAGGAACTCCCGAGCCTAGTGTTCACCTGGTCTTTAT 6133
Qy 1209 AsnGluThr-----IleProCysThrArgGluArgIleSerMetHisGlnAsp 1224
Db 6134 GATGGACTGAACATAAACCATTCGACTGCTCATTCACATTTTCTGTATCCAAAT 6193
Qy 1225 ThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyr 1244
Db 6194 GGAAC-----CTGTATATAAGAAGCATCGCTCCTTCAGTGAGGGGCACCTTAT 6241
Qy 1245 ThrLeuSerAlaLysAsnGluAlaGly----- 1253
Db 6242 GAGTGCAATTGCCACAGCTCCTCAGGCTCAGAGAGAGGGTAGTGATTTTACTGTGGAA 6301
Qy 1253 ----- 1253
Db 6302 GAGGGAGACAAATCCCGAGATAGAAACTGCCCTCAGAAATGGACTGAGGTGAATTTG 6361
Qy 1254 -----IleValSerCysThrAlaArgLeuAsp----- 1262
Db 6362 GGTGAGAAATTAATCTACTCACTGCTCAGCTACTGGGGATCCAAAGCCTAGAATAATCTGG 6421
Qy 1263 -----IleTyrAlaGlnTrpHisHisGlnIleProProPrometSer 1276
Db 6422 AGGTGCCATCAAGGCTGTCTCAGCAGTGG---CACAGAATGGCGAGCCGAATCCAC 6478
Qy 1277 ValArgProSerGlySerArg---TyrGlySerLeuThrLys 1290
Db 6479 GTCTACCCAAATGATGCTTGGTGGTGGTACGTGACCGGAAAAA 6523

Search completed: November 30, 2002, 22:52:06
Job time : 319 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 30, 2002, 19:09:50 ; Search time 2652 Seconds
(without alignments)
8061.102 Million cell updates/sec

Title: US-09-818-990B-2

Perfect score: 6890

Sequence: 1 MODSIEASTISQILLRESY.....MESTMVYSCSRSSVVEDEL 1320

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL-frame+ p2n model -DEV=xlh

-Q/cgn2_1/USPTO_spool/US09818990/runat_26112002_093405_23231/app_query.fasta_1.1479
-DB-EST -QFWT-fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=biosum62 -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1.1.1996 @runat_26112002_093405_23231 -NCFU=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1375.5	20.0	3865	11	AF077041	Homo sapi
2	1298.5	18.8	1285	11	AK010350	Mus muscu
3	965	14.0	614	14	BM968035	LM24HM007
4	910.5	13.2	901	14	BQ881261	AGENCOURT
5	908	13.2	595	13	BJ071900	BJ071900
6	887	12.9	943	13	BI691193	6033114585
7	877	12.7	868	12	BG827059	602749109
8	869.5	12.6	921	13	BI758769	60323796
9	858.5	12.5	717	14	BQ206762	UI-R-D21-
10	841	12.2	553	12	BF126187	601650450
11	831	12.1	828	13	BI526367	602925663
12	799	11.6	612	14	BM795050	K-EST0076
13	779	11.3	481	9	AA179599	zP49f09.f
14	768	11.1	654	10	BB548628	BB548628
15	758	11.0	772	13	BG923162	602824001
16	733	10.6	579	13	BJ041619	BJ041619
17	728	10.6	674	13	BI414869	602990942
18	711	10.3	1050	17	CNS05HM4	AL337765 Tetraodon
19	704	10.2	2209	11	AK017447	AK017447 Mus muscu
20	680	9.9	1101	17	CNS05AZS	AL329185 Tetraodon
21	679.5	9.9	874	12	BG423154	602450365
22	678.5	9.8	563	13	BJ069302	BJ069302
23	671.5	9.7	413	14	BQ559749	BQ559749 H4060802-
24	663	9.6	804	13	BI662601	603302549
25	662	9.6	506	9	AL709148	DRFZP686L
26	658	9.6	376	10	AW918754	AW918754 EST350058
27	652	9.5	503	17	AQ621045	HS_2221.B
28	646	9.4	444	13	BJ070056	BJ070056
29	643	9.3	594	13	BJ065434	BJ065434
30	639.5	9.3	969	14	BQ722262	AGENCOURT
31	638	9.3	804	14	BQ202926	UI-R-D01-
32	633	9.2	438	10	AW315222	12352 MAR
33	631.5	9.2	706	9	AI745576	WC34901.x
34	621.5	9.0	713	9	AA887587	OJ54H02.S
35	619	9.0	483	12	BE855373	ux51h11.Y
36	614.5	8.9	681	10	BB045059	BB045059
37	598	8.7	616	12	BF238982	601905239
38	597.5	8.7	558	9	AA128252	2130b11.f
39	580.5	8.4	670	10	BB486332	BB486332
40	575	8.3	531	9	AA446042	zw66b03.f
41	572.5	8.3	707	9	AL598867	DRFZP3130
42	572	8.3	501	10	BE448779	BE448779 ut83q05.Y
43	571.5	8.3	569	9	AA187765	AA187765 zP61a10.f
44	569	8.3	513	10	BB134094	BB134094
45	562	8.2	506	9	AA285584	vb83f11.f

ALIGNMENTS

RESULT 1	AF077041	Homo sapiens	3865 bp	mRNA	linear	HTC 22-MAY-2001
LOCUS	AF077041	Homo sapiens	SIH002 mRNA, complete cds.			
DEFINITION	AF077041	Homo sapiens	SIH002 mRNA, complete cds.			
ACCESSION	AF077041	Homo sapiens	SIH002 mRNA, complete cds.			
VERSION	AF077041.1	GI:4689129				
KEYWORDS	HTC.					
SOURCE	Homo sapiens.					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 3865)				
AUTHORS	Liu,T., Zhang,J., Ye,M., Zhang,Q., Fu,G., Zhou,J., Wu,J., Shen,Y., Yu,M., Chen,S., Mao,M. and Chen,Z.					
TITLE	Human SIH002 gene					


```

Db 370 AGAAGGCGCTCGCTTCGATCATCGGAGACAGTGGAGATGAAACAGAGCCATTTCAGGACGA 429
Qy 1070 PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeu 1089
Db 430 TTCTTCAGACTCACTTCCTGCGAGGCTTCCTGGAGACCTGACCGTTCAGGAGGCAAGCTC 489
Qy 1090 CysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsn 1109
Db 490 TGCAGGATGACTGCAAGTCACTGGGTACCAACCCAGATCTCAGCTGGCAACTAGAT 549
Qy 1110 GlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis 1129
Db 550 GGAAGGCCATAGCCGCCAGCAGTGCCTACAAGATGCTGGTCCGTGAGAATGGGGTCCAC 609
Qy 1130 SerLeuLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThr 1149
Db 610 TCCCTCATATTAGAGCAGTCACTCCCGGAGCGCGCATCTACACATGATTGGCCACC 669
Qy 1150 AsnLysThrGlyGlnAnSerPheSerLeuGluLeuSerValValAlaLysGluValLys 1169
Db 670 AACAGCAGCAGACAGACTCGTTTAACCTGGAGCTTGTGCTGCTAAGGAAGCACAC 729
Qy 1170 LysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProVal 1189
Db 730 AAGGCCCTGTGTTATTGAGAGAGCTACAGAACACGCGGGTGTGCTGATGGATACCCAGTG 789
Qy 1190 ArgLeuGluCysArgValIleGlyMetProProValPheTrpTrpLysLysAspAsn 1209
Db 790 CGCGTGAATGCCGTGCTCGGAGTGCCTCCAGCTCAGATATTTTGGAGAAGAAAT 849
Qy 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTrpAla 1229
Db 850 GAATCGCTCACTCACACCACTGAGCGAGTAGCATGACAGGATATCATGCGCTACATC 909
Qy 1230 CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTrpTrpLysLeuSerAlaLys 1249
Db 910 TGCGTGTCTATCCAGGAGGACCAAGAGAGACGCTGGTGGTACATGTGTGCGGCAAG 969
Qy 1250 AsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTrpAlaGlnTrpHisHis 1269
Db 970 AACGAACGAGCATGTGTCTGCACTGCGCGGCTGGATGTCTACACCCAGTGGACACAG 1029
Qy 1270 -----GlnIleProProMetSerValArgProSerGlySerArgTrpGlySerLeu 1287
Db 1030 CAGCCACAGACCAACCAAGCAAAAAAGTAGCGGCCCTCGGCCAGTCTGCGCAGCACTT 1089
Qy 1288 ThrSerLysGlyLeuAspIlePheSerAlaPheSerSerMetGluSerThrMetValTrp 1307
Db 1090 TCGGACCAAGGAGTAGACATCAAGCCGCTTTCACACT--GAAGCCAGCCCATCTCAC 1146
Qy 1308 SerCysSerArgSerValValGluSerAspGluLeu 1320
Db 1147 CTGACTTTGAACCGGCTTGGTAGAAGTGAAGACCTG 1185

RESULT 3
BM968035
LOCUS
DEFINITION
  LM24HW00787 Bos taurus LM-24-HW cDNA library Bos taurus cDNA clone
  LM-24-HW-007-87 (5'), mRNA sequence.
ACCESSION
  BM968035
VERSION
  BM968035.1 GI:19562450
KEYWORDS
  EST.
SOURCE
  cow.
  Bos taurus
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovine; Bos.
REFERENCE
  1 (bases 1 to 614)
  Yoon,D.H., Jang,Y.S., Kim,T.H., Park,E.W., Lee,H.K., Chung,E.R.,
  Sun,S.S. and Cheong,I.C.
  Gene Expression Profiling of the Bovine skeletal muscle
  Unpublished (2002)

```

```

COMMENT
Contact: Dr. Du-Hak Yoon
National Livestock Research Institute, RDA
564 Omoekchun-dong, Suwon, 441-350, Korea
Tel: 82 31 290 1593
Fax: 82 31 290 1792
Email: dhyoon@rda.go.kr
Insert Length: 614 Std Error: 0.00
Seq primer: CAGGAACAGCTATGAC
POLYA-No.
FEATURES
source
  1..614
    /organism="Bos taurus"
    /db_xref="taxon:9913"
    /clone="LM-24-HW-007-87 (5')"
    /clone_lib="Bos taurus LM-24-HW cDNA library"
    /sex="six males mixed"
    /tissue_type="longissimus dorsi"
    /cell_type="myocyte"
    /dev_stage="24 months old"
    /lab_host="XLI-BlueMRF strain"
    /note="Organ: skeletal muscle; Vector: Uni-ZAPXR; Site_1:
  EcORI; Site_2: Xho I"
  EcORI: 150 a 188 c 159 g 117 t
  BASE COUNT 150 a 188 c 159 g 117 t
  ORIGIN
Alignment Scores:
  Pred. No.: 7,01e-77 Length: 614
  Score: 965.00 Matches: 190
  Percent Similarity: 93.66% Conservative: 2
  Best Local Similarity: 92.68% Mismatches: 10
  Query Match: 14.01% Indels: 4
  DB: 14 Gaps: 2
  US-09-818-990B-2 (1-1320) x BM968035 (1-614)

```

```

Qy 1070 PhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeu 1089
Db 2 TTCTTCGAGCACATTCCTGCGAGCTCCTGGGATATGGTGGCTCAGAGGGCGGCTC 61
Qy 1090 CysArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsn 1109
Db 62 TGCGGCTGGAGCTGAAGTGAGTGGCTTACCGCCCCGGAGCTGACGTGGCTCTCAAT 121
Qy 1110 GlyGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHis 1129
Db 122 GGCAGCGCTGCTCCCGCACACCTCCCAAGATGCTGTCAGGAGACGGGAGTCCAC 181
Qy 1130 SerLeuLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThr 1149
Db 182 TCTGTGCTCATTTGACCGCTCACTCAACGCGACGCGGGACCTACACCTGCATTGCTACC 241
Qy 1150 AsnLysThrGlyGlnAnSerPheSerLeuGluLeuSerValValAlaLysGluValLys 1169
Db 242 AACAAAACCTGGCGAGAATTCCTTTAGCTTGGAGCTACTAGTAGCCCAAGAAAGTAAG 301
Qy 1170 LysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProVal 1189
Db 302 AAGCACCTCTGAATCCTCGAGAAACTCAGAACAGCGGTGTTCAGAAAGGCCACCCCGTG 361
Qy 1190 ArgLeuGluCysArgValIleGlyMetProProValPheTrpTrpLysLysAspAsn 1209
Db 362 AGACTGGAGTGGCGGGTGATAGCATCGCTCCCGCCCTGCTTCTACTGGAAGAAAGACAAC 421
Qy 1210 GluThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTrpAla 1229
Db 422 GAGACCATCCCTTTCACACAGAGAGAGATCAGCATGCGACGACACACTGGGTATGTC 481
Qy 1230 CysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTrpTrpLysLeuSerAlaLys 1249
Db 482 TGCCTCTCTATTCAGCCAGCAAAAAATCAGATGCTGGATGGTACACACTGTACGCAAG 541
Qy 1250 AsnGluAlaGlyIleValSerCysThr-AlaArgLeuAspIleTrpAlaGlnTrpHisHis 1269
  |||||||||||||||||||||||||||

```

```

Db 542 AC-GAAGCGGGCATCGTCTCTCACTGGCA---GCTGATATATACGCTATG---GATCA 594
QY 1269 sGlnIleProPro 1273
|::: |||||
Db 595 CAAACCCCCCA 607

RESULT 4
BQ881261
LOCUS
DEFINITION BQ881261 901 bp mRNA linear EST 16-AUG-2002
AGENCOURT.7978307 Lupski dorsal_root_ganglion Homo sapiens cDNA
clone IMAGE:6185000 5', mRNA sequence.
BQ881261
VERSION BQ881261.1 GI:22273269
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 901)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13575 row: b column: 09
High quality sequence stop: 610.

FEATURES
source
1..901
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6185000"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-GACTAGTCTGATCGAGCGCGCCCT(15)-3' and
5'-TCGACCCAGCGTCCG-3' and
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 260 a 232 c 223 g 186 t
ORIGIN

Alignment Scores:
Pred. No.: 1.07e-71 Length: 901
Score: 910.50 Matches: 169
Percent Similarity: 79.39% Conservative: 39
Best Local Similarity: 64.50% Mismatches: 50
Query Match: 13.21% Indels: 4
DB: 14 Gaps: 2

US-09-818-990B-2 (1-1320) x BQ881261 (1-901)
QY 1062 AsPlysGluProGlnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAsp 1081
::: |||||:::|||||
Db 34 GAAATGAACCAATTCAGGAGCATCTTCAGACCTCACTCTTCGAGGCTCTCGAGAT 93
::: |||||:::|||||
QY 1082 MetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeuProPro 1101
::: |||||:::|||||
Db 94 CTGACTGTTCAAGAAAGAAACTCTGCAGAAATGGACTGCAAGTCAAGTGGTTACCAACC 153

```

```

QY 1102 ProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMet 1121
|::: |||||:::|||||
Db 154 CCAGATCTAAGCTGGCAACTAGATGGAAGCCCGTACGCCCTCACAGTGTCTACAGATG 213
|::: |||||:::|||||
QY 1122 LeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAla 1141
|::: |||||:::|||||
Db 214 CTGGTGGTGGAGAACGGGTGCACCTCTCTGATCATAGAGCAGCTCACGTGATGATGCC 273
|::: |||||:::|||||
QY 1142 GlyThrTyrLysCysIleAlaThrAsnLysThrGlyClnAsnSerPheSerLeuGluLeu 1161
|::: |||||:::|||||
Db 274 GGCATCTACACATGTATAGCTACCAACCGAGCAGAGCAAGAACTCATTCAGCCTGGAGCTT 333
|::: |||||:::|||||
QY 1162 SerValValAlaLysGluValLysAlaProValIleLeuGluLysLeuGlnAsnCys 1181
|::: |||||:::|||||
Db 334 GTGGTTGCTGCTAAAGAACACACAAACCCCTGTGTTTATGAGAAGCTCCAAACACA 393
|::: |||||:::|||||
QY 1182 GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProPro 1201
|::: |||||:::|||||
Db 394 GGAGTTGCTGATGGGTACCCAGTGGCGTGAATGTCGTGTTATGGAGTGCACACACCT 453
|::: |||||:::|||||
QY 1202 ValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMet 1221
|::: |||||:::|||||
Db 454 CAGATATTTTGGAAAGAAAGAAATGAATCACTCACTCACAGCAGCTACCCGAGTGAGCATG 513
|::: |||||:::|||||
QY 1222 HisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysSerAspAla 1241
|::: |||||:::|||||
Db 514 CACGAGGACCAACACGGCTACATCTGCTGCTCATTCAGGGAGCCACAAAGAGATGCT 573
|::: |||||:::|||||
QY 1242 GlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeu 1261
|::: |||||:::|||||
Db 574 GGGTGTATATCTGTCAGCCAAAGAAATGAAGCAGGAGGATTCCTCTACTGCCAGGCTG 633
|::: |||||:::|||||
QY 1262 AsPileTyrAlaGlnTrpHisHis-----GlnIleProProPrometSerValArgPro 1279
|::: |||||:::|||||
Db 634 GACGTTTACACCCAGTGGCATGACGATGACAGCAGCAGCAAGCAAGCAAGTACGCCCC 693
|::: |||||:::|||||
QY 1280 SerGlySerArgTyrGlySerLeuThrSerGlyLeuAspIlePheSerAlaPheSer 1299
|::: |||||:::|||||
Db 694 TCAGGAGTGCCTATGTCAGCAGCTTCGGACCGAGGACTAGACATCAAGCAGCGTTCCCA 753
|::: |||||:::|||||
QY 1300 SerMetGluSerThrMetValTyrSerCysSerSerArgSerVal-ValGluSerAspG1 1319
|::: |||||:::|||||
Db 754 ACT---GAAGGCAACCCATCTCACTGACACTGAAATACTGCTTGGGTAGAAAGTGAAGA 810
|::: |||||:::|||||
QY 1319 uLeu 1320
|::: |||||
Db 811 ACTG 814

RESULT 5
BQ071900
LOCUS
DEFINITION BQ071900 595 bp mRNA linear EST 11-DEC-2001
laevis cDNA clone XL095nl8 5', mRNA sequence.
BQ071900
VERSION BQ071900.1 GI:17502089
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 595)
REFERENCE Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
AUTHORS y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855

```



```

||||| 800 GGACAGAACTCTTTCAGCGCTGGAGCTTGCTGCTAAAGAG 844
|||||
Db 800 603023796f1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194606 5',
EST 25-SEP-2001
RESULT 8 B1758769 921 bp mRNA linear EST 25-SEP-2001
LOCUS B1758769 921 bp mRNA linear EST 25-SEP-2001
DEFINITION 603023796f1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194606 5',
mRNA sequence.
ACCESSION B1758769
VERSION B1758769.1 GI:15750347
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 921)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cchapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M11486 row: 0 column: 23
High quality sequence start: 3
High quality sequence stop: 795.
Location/Qualifiers
1. .921
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5194606"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH MGC Library."
BASE COUNT 270 a 224 c 228 g 199 t
ORIGIN

Alignment Scores:
Pred. No.: 5,66e-68 Length: 921
Score: 869.50 Matches: 170
Percent Similarity: 73.67% Conservative: 51
Best Local Similarity: 56.67% Mismatches: 63
Query Match: 12.62% Indels: 16
DB: 13 Gaps: 5

US-09-818-990b-2 (1-1320) x B1758769 (1-921)

Qy 928 SerAspAspGluLeuGlnHisAspGluLeuProThrGlyLysCysIleAlaProIlePhe 947
Db 7 TCAGGTGATGAAGTTTCAGTGGAGATGCTGCTGGGAAATGGAATGGCACCATTCTTT 66
Qy 948 AspLysArgLeuLysHisPheArgValThrGluGlySerProValThrPheThrCysLys 967
Db 67 GAGATGAGCTGAACATTAAGAATCTTTGAGGAATGCCAGTAACCTTCACATGTAGA 126
Qy 968 IleValGlyIleProValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLys 987
Db 127 GTGGCTGGAATCCAAACCAAGATCTATTGGTTAAAGATGGGAACGACAGTCTCTCCA 186
Qy 988 ArgAsnGluHisCysLysMetArgArgGluGlyAspGlyThrCysSerLeuHisIleGlu 1007

```

```

Db 187 AAGAGTGATCACTACACCATTTCAAAGAGATCTCGATGGAGCTGCTCCTCCATACCACA 246
Qy 1008 SerThrThrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArg 1027
Db 247 GCCTCCACCCCTAGATGATGATGGGAATATATACAAATATGGCTGCAAAACCTTCAGGGCGC 306
Qy 1028 IleSerCysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArgLeuThrSer 1047
Db 307 ATCAGTTGTACTGCACGGCTAATGGTACAGGCTGTCAACCAAGAGGTCGAAGTCCCGCG 366
Qy 1048 AlaGlyGlnSerHis-----ArgGlyArgSerArgValGlnGluArg 1061
Db 367 TCTCCCTCAGGCCATCTCATGTCTCAGAAGGCTCGTTCTAGATCAAGGACAGTGGAGAC 426
Qy 1062 AspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAsp 1081
Db 427 GAAATGAACCAATTCAGGAGCGATCTTCAGACCTCACTTCTTCAGGGCTCCCTGGAGAT 486
Qy 1082 MetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProPro 1101
Db 487 CTGACTGTTCAAGAGGAAACTCTGCAGAATGGACTGCAAGTCAAGTGGGTACCAACC 546
Qy 1102 ProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMet 1121
Db 547 CCAGATCTAAGCTGCAACTAGATGGAAGCCCGTACGCCCTGACAGTGTCTACAAGATG 606
Qy 1122 LeuValArgGluThrGlyValHisSerLeuLeuAspProLeuThrGlnArgAspAla 1141
Db 607 CTGGTGGGTGAGAACGGGGTGCCTCTCTGATCATAGCCAGCCAGTCAAGTGGATGCC 666
Qy 1142 GlyThrTyrLysCysIle-AlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLe 1161
Db 667 GGCATCTACACATGTATAAGCTACCAACCGAGCAGGACAGAACTCATTCACCTGGAGCT 726
Qy 1161 uSerValValAlaLysGluValLysLysAlaProVal---IleLeuGluLysLeuGln-- 1179
Db 727 TGTGTAGTCTCTAAAGAAAGACACAAACCCCTGTGTATTATTGAGAAAGCTCCAAA 786
Qy 1180 -----AsnCysGlyValProGluGlyHisProValArg-LeuGluCysArgValIleG 1197
Db 787 ACACAGGAGTTGCTTGTATG-----CGGTACCCAGTGGGATGGAGTGTCCAGTTATTG 840
Qy 1197 lyMetProProValPhe---Tyr-TrpLysLysAspAsnGluThrIlePro 1213
Db 841 CGGAATGCCCCACCACACTTCGACTATTTTGGACGACAGACAGAACATGAATCAACACT 894
RESULT 9
BQ206762/c 717 bp mRNA linear EST 02-MAY-2002
LOCUS BQ206762
DEFINITION UI-R-DZ11-cnm-k-05-0-UI.s1 UI-R-DZ1 Rattus norvegicus cDNA clone
UI-R-DZ11-cnm-k-05-0-UI 3', mRNA sequence.
ACCESSION BQ206762
VERSION BQ206762.1 GI:20423222
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 717)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

```

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized cartilaginous tumor library cDNA library preparation: M.B. Soares lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES		Location/Qualifiers			
source					
1. .717		/organism="Rattus norvegicus"			
		/strain="Sprague-Dawley"			
		/db_xref="taxon:10116"			
		/clone="UI-R-D21-cm-k-05-0-UI"			
		/clone_lib="UI-R-D21"			
		/tissue_type="Chondrosarcoma"			
		/dev_stage="37 days"			
		/lab_host="DH10B (Life Technologies)"			
		/notes="Organ: Spine; Vector: pT7n3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; UI-R-D21 is a normalized cDNA library containing the following tissue(s): Swarm Rat Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7n3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CATCTTGTA. The Rat cartilaginous tumor tissue was provided by Dr Jeff Stevens at the University of Iowa.			
		TAG_LIB=UI-R-D21			
		TAG_TISSUE=cartilaginous tumor			
		TAG_SEQ=CATTCTGTA			
BASE COUNT	129 a 188 c 203 g 196 t	1 others			
ORIGIN					
Alignment Scores:					
Pred. No.:	3.8e-67	Length:	717		
Score:	858.50	Matches:	154		
Percent Similarity:	79.15%	Conservative:	32		
Best Local Similarity:	65.53%	Mismatches:	42		
Query Match:	12.46%	Indels:	7		
DB:	14	Gaps:	2		
US-09-818-990b-2 (1-1320) x BQ206762 (1-717)					
Qy	1049	GlycInSerHisArgGlyArgSerArgValGlnGluArgasp-----LysGlu	1064		
Db	712	GGTCATCTCTATGCAGCAAGCCCGCTCTCGATCAGCGACATGGAGATGAATGAG	653		
Qy	1065	ProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetValala	1084		
Db	652	CCCATTCAGGAGCGATTCTTCAGACCTCACTTCTCGAGGCTCTGGAGACCTCAGCGTT	593		
Qy	1085	HisGluGlyArgLeuGlyCysArgLeuAspCysLysValSerGlyLeuProProProGluLeu	1104		
Db	592	CAGGAAGCAAGCTCTCGAGGATGACTGCAAGGTCAGTGGATTACCAACCCAGATCTC	533		
Qy	1105	ThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuValArg	1124		
Db	532	AGCTGGCAACTAGACGGAAGCCCATCCGCCCTGACAGTGTCTACAGATGCTGTCGGT	473		
Qy	1125	GluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyr	1144		
Db	472	GAGATGGGTGGCACTCCCTCATTTATAGCCAGTCAGTCCCGGAGCGAGGATCTAC	413		
Qy	1145	LysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValVal	1164		

Db	412	ACGTGCATCGCCACCAACAGCAGCAGCGCCGATCATTTAACCTGGAGCTTGCTGCT	353
Qy	1165	AlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValPro	1184
Db	352	GCTAAAGACACACAGGCCCTGTGTATTCGAGAAGCTGCAAAACACGGGGTGGCC	293
Qy	1185	GluGlyHisProValArgLeuGluCysArgValIleGlyMetProProValPheTyr	1204
Db	292	GATGGGTACCCCGTCGCTGGAATGCGCGTTTCGGGGTGGCCGACCTCAGATATTC	233
Qy	1205	TrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMetHisGlnAsp	1224
Db	232	TGGAAGAAAGAAATGAATCGCTCACTCAGCAGCACTGATCGAGTCAGCATCCAGGAT	173
Qy	1225	ThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGlyThrTyr	1244
Db	172	ATCATGGCTACATCTGCTGCTCATTACAGGAGACACAAAGAGACGCTGGGTGGTAT	113
Qy	1245	ThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyr	1264
Db	112	ACGTGTCGCCCAAGAGCAGGAGCATCGTGTCTGTACTGCCAGGCTGGATGTCTAC	53
Qy	1265	AlaGlnTrpHisHisGlnIleProProMetSerValArgPro	1279
Db	52	ACCAGTGGCATCAGCAG-----CCACAGACCACCAAGCCA	17
RESULT 10			
BF126187			
LOCUS	BF126187	553 bp	mRNA linear EST 24-OCT-2000
DEFINITION	601650450F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:3934045 5', mRNA sequence.		
ACCESSION	BF126187		
VERSION	BF126187.1	GI:10965227	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 553)		
TITLE	NIH-MGC http://mgi.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: coqbs-r@mail.nih.gov		
	Tissue Procurement: CLONETECH Laboratories, Inc.		
	cDNA Library Preparation: CLONETECH Laboratories, Inc.		
	DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Plate: LUCM774 row: d column: 14		
	High quality sequence stop: 542.		
FEATURES	Location/Qualifiers		
source	1. .553		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3934045"		
	/clone_lib="NIH_MGC_76"		
	/lab_host="DH10B (TI phage-resistant)"		
	/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCCGAGCGGCGGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."		
BASE COUNT	143 a 138 c 153 g 119 t		
ORIGIN			

Alignment Scores:
Pred. NO.: 9.71e-66 Length: 553
Score: 841.00 Matches: 158
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.21% Indels: 0
DB: 12 Gaps: 0

US-09-818-990B-2 (1-1320) x BF126187 (1-553)

```
Qy 1163 ValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGly 1182
Db 2 GTAGTAGCAAGAGGTGAAGAAAGACCTGTGCTGGGAAACTACAGAACTGCGGT 61
Qy 1183 ValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProVal 1202
Db 62 GTTCCCGAAGCCACCCGCTGAGCTGGAGTGGCGGTGATAGGCATGCCCCACCTGTG 121
Qy 1203 PheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMethis 1222
Db 122 TTCTACTGGAAGAAAGACAATGAGACCATCCTTGCACACAGAGAGGATCAGTATGCAC 181
Qy 1223 GlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAlaGly 1242
Db 182 CAGGACAAACAGGGTATGCTGCTTCTCATTCAGCCAGCCAAAGAAATCAGACGCTGA 241
Qy 1243 TrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 1262
Db 242 TGGTACAGTTGTGAGCCAAAGAAAGAGCCGCGCATCTGTCGACATGCCAGGCTGGAT 301
Qy 1263 IleTyrAlaGlnTrpHisHisGlnIleProProMetSerValArgProSerGlySer 1282
Db 302 ATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGTCTGTCGCGCCAGTGGCAGT 361
Qy 1283 ArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerMetGlu 1302
Db 362 CGCTACGGATCTTCACACAGTAAGAGGACTTACATATTTCTGCGCTTTCTCCATGGA 421
Qy 1303 SerThrMetValTyrSerCysSerSerArgSerValValGluSerAspGluLeu 1320
Db 422 AGCAGATGGTGATTTCATGCTCTCTCGGAGTGTAGTGAGAGTGATGAATT 475
```

RESULT 11
BI526367 602925663F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5058231 5',
LOCUS mRNA sequence.
DEFINITION NIH-MGC http://mgi.nci.nih.gov/.
ACCESSION BI526367.1 GI:15351159
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 828)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1157 row: m column: 16
High quality sequence start: 25
High quality sequence stop: 826.
FEATURES
source location/Qualifiers
1. .828

```
/organism="Mus musculus"  
/strain="CZECH II"  
/db_xref="taxon:10090"  
/clone="IMAGE:5058231"  
/clone_lib="NCI_CGAP_Lu33"  
/tissue_type="pooled lung tumors"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: p7T3D-Pac (Pharmacia) with a  
modified polylinker; Site:1: NotI; Site:2: EcoRI; 1st  
strand cDNA was prepared from mRNA obtained from pooled  
lung tumors with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCTCTGTTTTTTTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified p7T3 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 211 a 228 c 223 g 166 t  
ORIGIN
```

Alignment Scores:

Pred. No.: 1.45e-64 Length: 828
Score: 831.00 Matches: 159
Percent Similarity: 76.21% Conservative: 46
Best Local Similarity: 59.11% Mismatches: 53
Query Match: 12.06% Indels: 11
DB: 13 Gaps: 2

US-09-818-990B-2 (1-1320) x BI526367 (1-828)

```
Qy 931 GluIleGlnHisAspGluIleProThrGlyLysCys-IleAlaProIlePheAspLysar 950
Db 30 AAGTGCAGGATCCAGATGTGCTGTGGAGAACGCCAACACCTCCCTCTTTGAGATGAA 89
Qy 950 gLeu-LysHisPheArgValThrGluGlySerProValThrPheThrCysLysIleValg 970
Db 90 GCTGGAACACATACCAAGATCTTGGTTAAAGATGGAGACAGATTTCTCCGAGAGCG 209
Qy 970 lyIleProValProLysValTyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnG 990
Db 150 GGAATCCAAAGCCAAAGATCTATTGGTTAAAGATGGAGACAGATTTCTCCGAGAGCG 209
Qy 990 luHisCysLysMetArgGluGlyAspGlyThrCysSerLeuHisIleGluSerThr 1010
Db 210 ATCACTACCATTCAGAGACCTTGTATGGAGCTGCTCTCTCCACACACCGCTCTA 269
Qy 1010 hrSerAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerC 1030
Db 270 CCCTAGACGAGATGGGAACATACACCATCATGTGCTGCCAACCCCTCAGGGTCGCTCAGTT 329
Qy 1030 ysSerGlyHisLeuMetValGlnSerLeuProIleArgSerArg-----LeuThrSerA 1048
Db 330 GTACAGGAAGCTAATGGTACAGGCTGTCAACCAAGAGCGCGGAGTCCCGCTCTCCCT 389
Qy 1048 laGlyGlnSerHisArgGlyArgSerValGlnGluArgAsp-----LysG 1064
Db 390 CAGGCCATCTCATGCCAAGAGCGCTGCTCTCATCAGCGGACAGTGGAGATGAAACG 449
Qy 1064 luProLeuGlnGluArgPheArgProHisPheLeuGlnAlaProGlyAspMetVala 1084
Db 450 AGCCCATTCAGAGCGGATTTCTCAGACCTCACTTCTCTGACGGCTCCTGGAGACCTGACCG 509
Qy 1084 laHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluL 1104
Db 510 TTCAGGAAGGCAAGCTCTGCAGGATGGACTCAAGGTGAGTGGGTACCAACCCAGATC 569
Qy 1104 euThrTrpLeuAsnGlyGlnProValLeuProAspAlaSerHisLysMetLeuVala 1124
Db 570 TCAGCTGGCAACTAGATGGAAAGCCCATACGCCCGCAGAGTGTCTCACAAGATGTGTGCC 629
Qy 1124 rgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAspIleThr 1144
Db 630 GTGGAATGGGGTCCCTCCCTCATATTAGAGCAGTACGCTCCCGGACCGCGGCATCT 689
```

US-09-818-990B-2 (1-1320) x BM795050 (1-612)

was cleaved with BamHI and XhoI. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT
ORIGIN

157 a 192 c 177 g 128 t

Alignment Scores:

Pred. No.: 5,02e-59 Length: 654
Score: 768.00 Matches: 140
Percent Similarity: 79.36% Conservative: 33
Best Local Similarity: 64.22% Mismatches: 39
Query Match: 11.15% Indels: 6
DB: 10 Gaps: 2

US-09-818-990B-2 (1-1320) x BB548628 (1-654)

QY 994 MetArgGluGluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAsp 1013

Db 1 ATTCAGAGAGACCTTGATGGAGACCTGCTCTCCACACACCGCCCTACCCCTAGAGAC 60

QY 1014 AspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHis 1033

Db 61 GATGGGAACCTACACCATCATGCTGCCAACCCCTCAGGTCGCTCAGTGTCTAGGAGG 120

QY 1034 LeuMetValGlnSerLeuProIleArgSerArg-----LeuThrSerAlaGlnSer 1051

Db 121 CTAATGGTACAGCTGTCAACCAAGAGCGCGAGTCCCGCTCTCCTCAGGCCATCCT 180

QY 1052 HisArgGlyArgSerArgValGlnGluArgAsp-----LysGluProLeuGln 1067

Db 181 CATGCCAAGAGCGCTCGCTCTCGATCAGCGGAGTGGAGATGAAACGAGCCCATCTAG 240

QY 1068 GluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGly 1087

Db 241 GACGGATTCTTACAGACCTCACTTCTCGAGCTCTCTGGAGACCTGACCGTTCAGGAGGC 300

QY 1088 ArgLeuCysArgLeuAspCysLysValSerGlyLeuProProGluLeuLeuThrTrpLeu 1107

Db 301 AAGCTCTGCAGGATGGACTCAAGGTCAGTGGTGTACCAACCCAGATCTCAGCTGGCAA 360

QY 1108 LeuAsnGlnProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGly 1127

Db 361 CTAGATGGAAGCCCATACCCCGAGTCAGTGCTCACAAAGATGCTGGTCCGTGAGATGG 420

QY 1128 ValHisSerLeuLeuLeuAspProLeuThrGlnArgAspAlaGlyThrTyLysCysIle 1147

Db 421 GTCCACTCCCTCATATAGAGCCAGTCACCTCCCGGACCGCGCATCTACATGTATT 480

QY 1148 AlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGlu 1167

Db 481 GCCACCAACAGAGCAGGACAGAACTCGTTTAACCTGGAGCTTGTGTGCTGCTAAGGAA 540

QY 1168 ValLysLysAlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHis 1187

Db 541 GCACACAGCCCTCGTGTATGGAGAGCTACAGAACACGGGGTGTGTGATGGATAC 600

QY 1188 ProValArgLeuGluCysArgValIleGlyMetProProValPheTyTrp 1205

Db 601 CCAGTCGGCTGGAATGCGCTGCTCGGAGTGCCGCCACCTCAGATATTTTG 654

RESULT 15 BG923162 772 bp mRNA linear EST 05-JUN-2001

LOCUS 602824001F1 NCI_CGAP_Mam6 Mus musculus cdna clone IMAGE:4952713 5',

DEFINITION mRNA sequence.

ACCESSION BG923162

VERSION BG923162.1 GI:14303638

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 772)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE
JOURNAL

COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, ph.D.
Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10911 row: a column: 02

High quality sequence stop: 765.

FEATURES

source

1. .772
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4952713"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
Providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 198 a 229 c 202 g 143 t

ORIGIN

Alignment Scores:

Pred. No.: 5,2e-58 Length: 772
Score: 758.00 Matches: 140
Percent Similarity: 80.48% Conservative: 29
Best Local Similarity: 66.67% Mismatches: 38
Query Match: 11.00% Indels: 3
DB: 13 Gaps: 1

US-09-818-990B-2 (1-1320) x BG923162 (1-772)

QY 1092 LeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGln 1111

Db 1 ATGGACTGCAAGGTCACTGGTGTACCAACCCAGATCTCAGCTGGCACTAGATGGAAG 60

QY 1112 ProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeu 1131

Db 61 CCCATGCCCCGACAGTGTCTACAAAGATGCTGCTGCTGAGATGGGGTCCACTCCCTC 120

QY 1132 LeuIleAspProLeuThrGlnArgAspAlaGlyThrTyLysCysIleAlaThrAsnLys 1151

Db 121 ATTATAGCCAGTCAGCTCCCGGACCGCGGCATCTACATGTATGGCCACCAACAGA 180

QY 1152 ThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysAla 1171

Db 181 GCAGGACAGAACTCGTTTAACCTGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 240

QY 1172 ProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeu 1191

Db 241 CCTGTGTTTATGGAAGAGCTACAGAACACGGGGTGTCTGATGATACCCAGTCGGCTG 300

QY 1192 GluCysArgValIleGlyMetProProValPheTyTrpLysLysAspAsnGluThr 1211

Db 301 GAATGCGGTGCTCTCGGAGTGCCGCCACCTCAGATATTTTGGAGAAGAAATGAATCG 360

QY 1212 IleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyAlaCysLeu 1231

Db 361 CTCACCTCAGACACTGAGCGAGTAAAGCATGCACAGGATAATATGCTACATCTGCTG 420

QY 1232 LeuIleGlnProAlaLysLysSerAspAla-GlyTyTrpTyThrLeuSerAlaLysAsnG 1251

Db 421 CTCATCCAGGAGCCACAAAGGAGACGCTGGCTGGTGTACACTGTGTCCGCCAAGACGA 480

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 14:01:04 ; Search time 6626 Seconds
(without alignments)
17406.342 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 3963

Sequence: 1 atgcaagacagacagcataga.....tggagagtgtgaacttttaa 3963

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3504	88.4	5707	9	AF328296	AF328296 Homo sapi
2	3453	87.1	5804	9	HS805275	AL834247 Homo sapi
3	3249	82.0	5486	9	HS803309	AL832002 Homo sapi
4	2602	65.7	4683	9	HS803687	AL832379 Homo sapi
5	1483	37.4	2127	9	AK027343	AK027343 Homo sapi
6	904	22.8	160558	9	AC024258	AC024258 Homo sapi
7	904	22.8	175265	2	AC063964	AC063964 Homo sapi
8	904	22.8	180707	2	AC024460	AC024460 Homo sapi
9	413	10.4	11170	9	AL512429	AL512429 Human DNA
10	413	10.4	290625	2	AL442068	AL442068 Homo sapi
11	236	6.0	175265	2	AC063964	AC063964 Homo sapi
12	227	5.7	290625	2	AL442068	AL442068 Homo sapi
13	210	5.3	166869	9	AC016395	AC016395 Homo sapi
14	41	1.0	151002	2	AC114220	AC114220 Rattus no
15	38	1.0	151002	2	AC114220	AC114220 Rattus no
16	37	0.9	512	11	AF139038	AF139038 Ovis arie
17	26	0.7	42861	2	AC120168	AC120168 Mus muscu
18	24	0.6	3135	3	AY057969	AY057969 Girardia
19	23	0.6	71656	2	AC067914	AC067914 Homo sapi
20	23	0.6	128824	2	AC094957	AC094957 Rattus no
21	23	0.6	147603	9	HS025883	AL109919 Human DNA
22	23	0.6	173222	9	AC009127	AC009127 Homo sapi
23	23	0.6	194732	9	AB042031	AB042031 Homo sapi
24	23	0.6	195741	2	AC098273	AC098273 Rattus no
25	23	0.6	300000	9	AP002529	AP002529 Homo sapi
26	22	0.6	68743	2	AC113547	AC113547 Mus muscu
27	22	0.6	136493	2	AC125500	AC125500 Takifugu
28	22	0.6	170000	2	AC004394	AC004394 Homo sapi
29	22	0.6	170000	2	AC004579	AC004579 Homo sapi
30	22	0.6	174138	9	AC008871	AC008871 Homo sapi
31	22	0.6	174891	2	AC113847	AC113847 Rattus no
32	22	0.6	190000	2	AC004580	AC004580 Homo sapi
33	22	0.6	196426	2	AC127234	AC127234 Mus muscu
34	22	0.6	210515	9	AC097382	AC097382 Homo sapi
35	21	0.5	1580	10	AF205078	AF205078 Mus muscu
36	21	0.5	3147	10	BC027364	BC027364 Mus muscu
37	21	0.5	37443	9	U73627	U73627 Human Chrom
38	21	0.5	51761	2	AC068574	AC068574 Homo sapi
39	21	0.5	62726	2	AC084090	AC084090 Homo sapi
40	21	0.5	70406	2	AC103376	AC103376 Mus muscu
41	21	0.5	71530	2	AC116112	AC116112 Mus muscu
42	21	0.5	94851	9	AC004923	AC004923 Homo sapi
43	21	0.5	94911	8	AT114E10	AL138656 Arabidops
44	21	0.5	106922	2	AC127029	AC127029 Homo sapi
45	21	0.5	108816	2	AL670379	AL670379 Homo sapi
46	21	0.5	115279	9	AC112506	AC112506 Homo sapi
47	21	0.5	127188	9	AC010490	AC010490 Homo sapi
48	21	0.5	127539	9	CNS01DSN	AL121821 Human chr
49	21	0.5	132830	2	AC011051	AC011051 Homo sapi
50	21	0.5	138915	2	AP005246	AP005246 Oryza sat
51	21	0.5	142918	2	AF252828	AF252828 Homo sapi
52	21	0.5	146558	2	AC090753	AC090753 Homo sapi
53	21	0.5	152141	2	AC119787	AC119787 Rattus no
54	21	0.5	153241	2	AC109690	AC109690 Rattus no
55	21	0.5	156307	2	AP005248	AP005248 Oryza sat
56	21	0.5	162143	10	AL596283	AL596283 Mouse DNA
57	21	0.5	162473	9	AC005803	AC005803 Homo sapi
58	21	0.5	167875	9	AC083906	AC083906 Homo sapi
59	21	0.5	170656	2	AF216672	AF216672 Homo sapi
60	21	0.5	177654	2	AC025988	AC025988 Homo sapi
61	21	0.5	178229	2	AC009420	AC009420 Homo sapi
62	21	0.5	185045	2	AC126303	AC126303 Rattus no
63	21	0.5	185148	9	AC096641	AC096641 Homo sapi
64	21	0.5	186824	2	AC121881	AC121881 Mus muscu
65	21	0.5	200833	9	AL390738	AL390738 Human DNA

212	20	0.5	204834	9	AC009312	AC009312 Homo sapi	285	19	0.5	4347	9	AB023209	AB023209 Homo sapi
213	20	0.5	207086	10	AL596117	AL596117 Mouse DNA	286	19	0.5	4871	5	AF337035	AF337035 Danio rer
214	20	0.5	207804	2	AC011714	AC011714 Homo sapi	287	19	0.5	4950	10	AF494464	AF494464 Mus muscu
215	20	0.5	208488	2	AC034212	AC034212 Homo sapi	288	19	0.5	5773	9	AF464873	AF464873 Homo sapi
216	20	0.5	208587	2	AC079498	AC079498 Mus muscu	289	19	0.5	6314	6	A23832	A23832 H.sapiens m
217	20	0.5	209003	2	AC129205	AC129205 Mus muscu	290	19	0.5	6314	6	AR011657	AR011657 Sequence
218	20	0.5	209572	2	AC099615	AC099615 Mus muscu	291	19	0.5	6314	6	AX332837	AX332837 Sequence
219	20	0.5	211445	9	AC103794	AC103794 Homo sapi	292	19	0.5	6314	6	AX333748	AX333748 Sequence
220	20	0.5	216016	2	AC090293	AC090293 Mus muscu	293	19	0.5	6314	9	HUMKALL	M97252 Homo sapien
221	20	0.5	217926	2	AC127360	AC127360 Mus muscu	294	19	0.5	7000	10	AF458960	AF458960 Mus muscu
222	20	0.5	224790	2	AL683819	AL683819 Mus muscu	295	19	0.5	7354	1	AE007720	AE007720 Clostridi
223	20	0.5	225204	2	AL772138	AL772138 Mus muscu	296	19	0.5	8062	1	AF292097	AF292097 Leprospir
224	20	0.5	234777	2	AC125223	AC125223 Mus muscu	297	19	0.5	8157	2	AL831743	AL831743 Mus muscu
225	20	0.5	244931	2	AC079495	AC079495 Mus muscu	298	19	0.5	9990	6	AX201603	AX201603 Sequence
226	20	0.5	252278	2	AC090126	AC090126 Mus muscu	299	19	0.5	9990	6	AX201616	AX201616 Sequence
227	20	0.5	269081	2	AC068493	AC068493 Mus muscu	300	19	0.5	10005	1	AE013008	AE013008 Thermoana
228	20	0.5	310360	2	AC074172	AC074172 Mus muscu	301	19	0.5	10165	9	AE015538	AE015538 Homo sapi
229	20	0.5	340000	9	AP001178	AP001178 Homo sapi	302	19	0.5	10218	1	AE004725	AE004725 Pseudomon
230	19	0.5	102	6	A74401	A74401 Sequence 87	303	19	0.5	10379	1	AE006667	AE006667 Sulfolobu
231	19	0.5	102	6	A74401	A77380 Sequence 87	304	19	0.5	10379	1	AE001085	AE001085 Archaeogl
232	19	0.5	360	6	AR202904	AR202904 Sequence	305	19	0.5	10478	1	AE001084	AE001084 Archaeogl
233	19	0.5	360	6	AR208005	AR208005 Sequence	306	19	0.5	11019	1	AE001084	AE001084 Archaeogl
234	19	0.5	360	6	AX429878	AX429878 Sequence	307	19	0.5	11339	1	AE012794	AE012794 Chloroblu
235	19	0.5	636	8	FRU71015	U71015 Festuca rub	308	19	0.5	12240	1	AE010273	AE010273 Pyrococcu
236	19	0.5	738	14	AF499687	AF499687 Lymnaetria	309	19	0.5	13512	1	AE001045	AE001045 Archaeogl
237	19	0.5	956	8	AY086402	AY086402 Arabidops	310	19	0.5	14030	1	AE0011629	AE0011629 Xanthomon
238	19	0.5	1204	8	AY117314	AY117314 Arabidops	311	19	0.5	14728	9	AB0116243	AB0116243 Homo sapi
239	19	0.5	1228	6	AX431829	AX431829 Sequence	312	19	0.5	15330	3	CE2RSSE	AL845459 Caenorhab
240	19	0.5	1296	6	AF026396	AF026396 Saccharom	313	19	0.5	16170	1	AE000959	AE000959 Archaeogl
241	19	0.5	1300	8	AF026396	BC017832 Homo sapi	314	19	0.5	16234	1	AE001060	AE001060 Archaeogl
242	19	0.5	1365	8	AY064047	AY064047 Arabidops	315	19	0.5	17854	1	AF307053	AF307053 Thermococ
243	19	0.5	1401	9	AK000845	AK000845 Homo sapi	316	19	0.5	18001	1	AF307052	AF307052 Pyrococcu
244	19	0.5	1409	9	AK025843	AK025843 Homo sapi	317	19	0.5	18150	9	AL731570	AL731570 Human DNA
245	19	0.5	1419	6	AX065465	AX065465 Sequence	318	19	0.5	23457	2	AE008913	AE008913 Salmonell
246	19	0.5	1419	6	AF419594	AF419594 Arabidops	319	19	0.5	23987	2	AC118516	AC118516 Rattus no
247	19	0.5	1467	6	AX065463	AX065463 Sequence	320	19	0.5	28108	8	AP004470	AP004470 Lotus jap
248	19	0.5	1514	9	BC017832	BC017832 Homo sapi	321	19	0.5	28437	9	AL606483	AL606483 Human DNA
249	19	0.5	1611	6	AR164079	AR164079 Sequence	322	19	0.5	31462	9	HSCAMHCA	U22056 Homo sapien
250	19	0.5	1611	9	HUMERK2A	M84489 Human extra	323	19	0.5	32365	3	U12964	U12964 Caenorhabdi
251	19	0.5	1650	9	BC019097	BC019097 Homo sapi	324	19	0.5	32026	9	AF156673	AF156673 Homo sapi
252	19	0.5	1663	3	DTNRASPK	X75310 D.tigrina s	325	19	0.5	33110	9	AL354671	AL354671 Human DNA
253	19	0.5	1664	3	BC013867	BC013867 Homo sapi	326	19	0.5	38418	10	MUSPNPA	U29186 Mus musculu
254	19	0.5	1846	9	AB062418	AB062418 Homo sapi	327	19	0.5	39760	10	U29187	U29187 Mus musculu
255	19	0.5	1905	1	CORACEA	L28760 Corynebacte	328	19	0.5	39854	9	AC004794	AC004794 Homo sapi
256	19	0.5	1939	6	HUMKERP1	M22927 Human kerat	329	19	0.5	39971	3	U13875	U13875 Caenorhabdi
257	19	0.5	1977	6	AX074149	AX074149 Sequence	330	19	0.5	40594	3	CER03A10	269793 Caenorhabdi
258	19	0.5	1986	6	AX179749	AX179749 Sequence	331	19	0.5	41280	9	AL589906	AL589906 Human DNA
259	19	0.5	1988	9	HS420421	AJ420421 Homo sapi	332	19	0.5	42170	2	AC104252	AC104252 Homo sapi
260	19	0.5	2024	10	BC032265	BC032265 Mus muscu	333	19	0.5	44323	3	AF098992	AF098992 Caenorhab
261	19	0.5	2075	6	E04376	E04376 DNA sequenc	334	19	0.5	49574	2	AC098338	AC098338 Rattus no
262	19	0.5	2133	9	AB062994	AB062994 Macaca fa	335	19	0.5	50861	2	AC079265	AC079265 Mus muscu
263	19	0.5	2135	6	I13693	I13693 Sequence 3	336	19	0.5	51827	2	AC101058	AC101058 Mus muscu
264	19	0.5	2135	6	I86191	I86191 Sequence 3	337	19	0.5	53993	2	AC100666	AC100666 Mus muscu
265	19	0.5	2162	9	AK022915	AK022915 Homo sapi	338	19	0.5	54343	2	AC121530	AC121530 Mus muscu
266	19	0.5	2188	8	PEAORF	L49510 Pisum sativ	339	19	0.5	55625	2	AC068382	AC068382 Homo sapi
267	19	0.5	2192	9	AK074246	AK074246 Homo sapi	340	19	0.5	58150	9	AC018918	AC018918 Homo sapi
268	19	0.5	2193	9	S60085S1	S60085 H.sapiens 4	341	19	0.5	60928	2	AC013830	AC013830 Homo sapi
269	19	0.5	2351	9	HS41KDAP	HS41KDAP	342	19	0.5	60928	2	AC013830	AC013830 Homo sapi
270	19	0.5	2386	1	AF317076	AF317076 Synchococ	343	19	0.5	61377	2	AC080054	AC080054 Homo sapi
271	19	0.5	2427	1	CGACEA	X75504 C.glutamicu	344	19	0.5	61559	2	AC125439	AC125439 Homo sapi
272	19	0.5	2494	9	AF231021	AF231021 Homo sapi	345	19	0.5	61720	2	AC131033	AC131033 Mus muscu
273	19	0.5	2509	9	HS40KDAP	HS40KDAP	346	19	0.5	62047	2	AC117672	AC117672 Mus muscu
274	19	0.5	2562	9	AK074255	AK074255 Homo sapi	347	19	0.5	62469	2	AC100670	AC100670 Homo sapi
275	19	0.5	2562	9	AK074255	AK074255 Homo sapi	348	19	0.5	62833	2	AC107942	AC107942 Homo sapi
276	19	0.5	2791	9	HS40KDAP	HS40KDAP	349	19	0.5	62979	2	AC097783	AC097783 Rattus no
277	19	0.5	2846	9	AK090869	AK090869 Homo sapi	350	19	0.5	63974	2	AC102816	AC102816 Mus muscu
278	19	0.5	2920	9	AK095512	AK095512 Homo sapi	351	19	0.5	64011	2	AC107683	AC107683 Mus muscu
279	19	0.5	3106	3	ASATPASE	U7513 A.francisca	352	19	0.5	64869	9	AL125429	AL125429 Homo sapi
280	19	0.5	3332	9	AK056654	AK056654 Homo sapi	353	19	0.5	65096	2	AL590648	AL590648 Human DNA
281	19	0.5	3546	9	HS40KDAP	HS40KDAP	354	19	0.5	65237	2	AC105208	AC105208 Homo sapi
282	19	0.5	3757	9	AF151909	AF151909 Homo sapi	355	19	0.5	65241	2	AC099968	AC099968 Mus muscu
283	19	0.5	4094	9	HSKALIG	X60299 H.sapiens K	356	19	0.5	65262	2	AC115710	AC115710 Mus muscu
284	19	0.5	4208	5	AF223970	AF223970 Gallus ga	357	19	0.5	65666	2	AC100486	AC100486 Mus muscu
			4325	10	RNU22952	U22952 Rattus norv				65874	2	AC130495	AC130495 Mus muscu

c 358	19	0.5	66442	2	AC100287	AC100287 Mus muscu	c 431	19	0.5	112303	9	AC093836	AC093836 Homo sapi
c 359	19	0.5	67523	2	AC104356	AC104356 Homo sapi	c 432	19	0.5	112930	2	AC125820	AC125820 Rattus no
c 360	19	0.5	67662	2	AC119898	AC119898 Mus muscu	c 433	19	0.5	113168	9	HSJ779E11	AL121675 Human DNA
c 361	19	0.5	67674	2	AC101032	AC101032 Mus muscu	434	19	0.5	114734	2	AC097595	AC097595 Rattus no
c 362	19	0.5	68027	2	AC124261	AC124261 Homo sapi	435	19	0.5	114800	2	AC123567	AC123567 Homo sapi
c 363	19	0.5	68111	2	AC110265	AC110265 Mus muscu	436	19	0.5	115641	8	F8K4	AC004392 Arabidops
c 364	19	0.5	68620	9	AL133293	AL133293 Human DNA	437	19	0.5	116569	2	AC102385	AC102385 Mus muscu
c 365	19	0.5	68762	2	AC120411	AC120411 Mus muscu	438	19	0.5	117217	9	AP002076	AP002076 Homo sapi
c 366	19	0.5	68720	2	AC120411	AC120411 Mus muscu	439	19	0.5	117954	9	HS409010	AL031256 Human DNA
c 367	19	0.5	68731	3	AC087415	AC087415 Caenorhab	c 440	19	0.5	118510	9	AL731543	AL731543 Human DNA
c 368	19	0.5	69133	2	AC110508	AC110508 Mus muscu	441	19	0.5	118716	9	AC116654	AC116654 Homo sapi
c 369	19	0.5	69133	2	AC110508	AC110508 Mus muscu	c 442	19	0.5	118831	9	HS989H11	283851 Human DNA s
c 370	19	0.5	69847	9	AC004845	AC004845 Homo sapi	c 443	19	0.5	119631	9	AC008888	AC008888 Homo sapi
c 371	19	0.5	70136	2	AC012699	AC012699 Drosophil	c 444	19	0.5	120079	2	AC096153	AC096153 Rattus no
c 372	19	0.5	70496	2	AC122555	AC122555 Mus muscu	c 445	19	0.5	121125	2	AC115312	AC115312 Rattus no
c 373	19	0.5	70949	2	AC123670	AC123670 Mus muscu	446	19	0.5	121492	8	OSJN00104	AL606653 Oryza sat
c 374	19	0.5	70953	2	AC118045	AC118045 Mus muscu	c 447	19	0.5	121889	2	AC121406	AC121406 Rattus no
c 375	19	0.5	71068	2	AC109266	AC109266 Mus muscu	448	19	0.5	122697	9	AC006062	AC006062 Homo sapi
c 376	19	0.5	72014	2	AC100639	AC100639 Mus muscu	449	19	0.5	122823	9	AC005996	AC005996 Homo sapi
c 377	19	0.5	73259	2	AC084006	AC084006 Homo sapi	c 450	19	0.5	123192	2	AP000589	AP000589 Homo sapi
c 378	19	0.5	74362	2	AC119211	AC119211 Mus muscu	c 451	19	0.5	123236	9	HS0J98E18	AL049694 Human DNA
c 379	19	0.5	75111	2	AC106330	AC106330 Rattus no	c 452	19	0.5	124829	9	AC097109	AC097109 Homo sapi
c 380	19	0.5	75733	2	AC006412	AC006412 Homo sapi	c 453	19	0.5	125196	8	OSJN00203	AL663003 Oryza sat
c 381	19	0.5	75994	9	AC105417	AC105417 Homo sapi	454	19	0.5	125962	2	AC084271	AC084271 Rattus no
c 382	19	0.5	76540	9	AC003678	AC003678 Homo sapi	c 455	19	0.5	125962	2	AC084271	AC084271 Rattus no
c 383	19	0.5	77017	9	AL139138	AL139138 Human DNA	c 456	19	0.5	127524	2	AL807806	AL807806 Mus muscu
c 384	19	0.5	77738	9	HS1195D24	AL121892 Human DNA	457	19	0.5	127714	8	OSJN00273	AL731630 Oryza sat
c 385	19	0.5	79586	9	AC005600	AC005600 Homo sapi	c 458	19	0.5	127989	9	AC010908	AC010908 Homo sapi
c 386	19	0.5	79590	2	AC010457	AC010457 Homo sapi	c 459	19	0.5	129109	9	AL390239	AL390239 Human DNA
c 387	19	0.5	81436	2	AC023552	AC023552 Homo sapi	c 460	19	0.5	129240	9	AC083826	AC083826 Homo sapi
c 388	19	0.5	81952	9	AC110801	AC110801 Homo sapi	461	19	0.5	129360	8	OSJN00108	AL606659 Oryza sat
c 389	19	0.5	82601	9	AL136971	AL136971 Human DNA	c 462	19	0.5	129923	9	AL445204	AL445204 Human DNA
c 390	19	0.5	83339	8	AP001311	AP001311 Arabidops	c 463	19	0.5	130407	2	AC084354	AC084354 Homo sapi
c 391	19	0.5	83661	9	AC011343	AC011343 Homo sapi	c 464	19	0.5	130638	2	AC105686	AC105686 Rattus no
c 392	19	0.5	83894	2	AC128300	AC128300 Rattus no	c 465	19	0.5	130753	9	HS510L9	AL022098 Homo sapi
c 393	19	0.5	84588	2	AC114832	AC114832 Rattus no	466	19	0.5	130851	2	AC111945	AC111945 Rattus no
c 394	19	0.5	85440	2	AC034264	AC034264 Mus muscu	467	19	0.5	131989	2	AC107545	AC107545 Rattus no
c 395	19	0.5	86064	2	AL391554	AL391554 Homo sapi	c 468	19	0.5	132324	2	AC004737	AC004737 Homo sapi
c 396	19	0.5	86519	9	AP001860	AP001860 Homo sapi	c 469	19	0.5	133508	9	AF241728	AF241728 Homo sapi
c 397	19	0.5	87248	9	AC114763	AC114763 Homo sapi	c 470	19	0.5	134102	8	AC074196	AC074196 Oryza sat
c 398	19	0.5	87978	9	AC003971	AC003971 Human BAC	c 471	19	0.5	134411	9	AL512662	AL512662 Human DNA
c 399	19	0.5	88037	1	AL627284	AL627284 Salmonell	472	19	0.5	135161	2	AL357494	AL357494 Homo sapi
c 400	19	0.5	88454	9	AC011412	AC011412 Homo sapi	c 473	19	0.5	135340	2	AC126352	AC126352 Homo sapi
c 401	19	0.5	88508	9	AC129729	AC129729 Homo sapi	c 474	19	0.5	135571	9	HS340N1	298257 Human DNA s
c 402	19	0.5	89208	9	AC114752	AC114752 Homo sapi	c 475	19	0.5	135631	2	AL512649	AL512649 Human DNA
c 403	19	0.5	89479	8	AC006932	AC006932 Genomic s	c 476	19	0.5	136357	9	AL161942	AL161942 Homo sapi
c 404	19	0.5	94911	8	ATT14E10	AL138656 Arabidops	c 477	19	0.5	136502	9	HS705F19	AL035415 Human DNA
c 405	19	0.5	95310	8	ATAC011437	AC011437 Arabidops	478	19	0.5	136836	2	AC130876	AC130876 Rattus no
c 406	19	0.5	95445	2	AC016175	AC016175 Homo sapi	479	19	0.5	136932	9	AC068035	AC068035 Homo sapi
c 407	19	0.5	96500	8	AP001243	AP001243 Homo sapi	480	19	0.5	137376	9	AC112205	AC112205 Homo sapi
c 408	19	0.5	98506	8	ATT2L20	AL592312 Arabidops	481	19	0.5	137433	2	AC108642	AC108642 Rattus no
c 409	19	0.5	98966	2	AC120510	AC120510 Gallus ga	482	19	0.5	137693	9	AC003689	AC003689 Homo sapi
c 410	19	0.5	99033	2	AC123261	AC123261 Rattus no	c 483	19	0.5	137724	8	AC122147	AC122147 Oryza sat
c 411	19	0.5	99878	2	AC087108	AC087108 Homo sapi	484	19	0.5	138224	9	AC096562	AC096562 Homo sapi
c 412	19	0.5	102445	2	AC114867	AC114867 Rattus no	485	19	0.5	139360	2	AC073500	AC073500 Homo sapi
c 413	19	0.5	102733	9	AC108865	AC108865 Homo sapi	486	19	0.5	140092	9	AL445591	AL445591 Human DNA
c 414	19	0.5	104457	2	AC098577	AC098577 Rattus no	c 487	19	0.5	140098	2	AC024510	AC024510 Homo sapi
c 415	19	0.5	105149	9	AP000484	AP000484 Homo sapi	c 488	19	0.5	140207	2	AC128666	AC128666 Mus muscu
c 416	19	0.5	105841	8	AC092697	AC092697 Oryza sat	489	19	0.5	140845	2	AC124351	AC124351 Mus muscu
c 417	19	0.5	106179	9	HS790B6	AL031677 Human DNA	490	19	0.5	140908	2	AC131470	AC131470 Human DNA
c 418	19	0.5	108554	2	AC123336	AC123336 Rattus no	491	19	0.5	141452	9	AL137880	AL137880 Human DNA
c 419	19	0.5	108967	9	AL353138	AL353138 Human DNA	c 492	19	0.5	142441	2	HSJ344H20	AL132710 Homo sapi
c 420	19	0.5	109071	9	AC099482	AC099482 Homo sapi	c 493	19	0.5	142542	2	AC127828	AC127828 Rattus no
c 421	19	0.5	109616	9	AC008723	AC008723 Homo sapi	c 494	19	0.5	142969	2	AC025840	AC025840 Homo sapi
c 422	19	0.5	110000	2	AC097795_1	Continuation (2 of	495	19	0.5	143176	2	AC129122	AC129122 Rattus no
c 423	19	0.5	110000	2	AC013623_2	Continuation (3 of	c 496	19	0.5	143176	2	AC129122	AC129122 Rattus no
c 424	19	0.5	110000	2	LMFLCHR18_08	Continuation (9 of	c 497	19	0.5	143282	2	AC123303	AC123303 Rattus no
c 425	19	0.5	110000	9	AF491780_05	Continuation (6 of	c 498	19	0.5	143370	2	AC027587	AC027587 Homo sapi
c 426	19	0.5	110000	9	AF491780_06	Continuation (7 of	499	19	0.5	143413	2	AC025979	AC025979 Homo sapi
c 427	19	0.5	110000	10	AE014175_1	Continuation (2 of	c 500	19	0.5	143489	2	AL161637	AL161637 Homo sapi
c 428	19	0.5	110104	9	AC011348	AC011348 Homo sapi	c 501	19	0.5	143885	2	AC119339	AC119339 Rattus no
c 429	19	0.5	110892	2	AP004678	AP004678 Oryza sat	502	19	0.5	144177	9	AL513487	AL513487 Human DNA
c 430	19	0.5	111690	2	AC130954	AC130954 Rattus no	503	19	0.5	144487	2	AC119366	AC119366 Rattus no

650	19	0.5	173843	2	AC096018	AC096018 Rattus no	723	19	0.5	185132	9	CNS01DUB	AL133238 Human chr
651	19	0.5	173997	2	AC098949	AC098949 Rattus no	724	19	0.5	185165	9	AC008732	AC008732 Homo sapi
652	19	0.5	174082	2	AC007379	AC007379 Homo sapi	c 725	19	0.5	185207	2	AC084350	AC084350 Homo sapi
653	19	0.5	174248	2	AC113654	Rattus no	726	19	0.5	185655	10	MMU296304	AJ296304 Mus muscu
654	19	0.5	174318	9	CNS01RGC	AL159141 Human chr	727	19	0.5	185828	2	AC122815	AC122815 Mus muscu
655	19	0.5	174399	9	AC104576	AC104576 Homo sapi	c 728	19	0.5	186269	2	AC095430	AC095430 Rattus no
656	19	0.5	174879	2	AC090602	AC090602 Homo sapi	729	19	0.5	186760	2	AC120457	AC120457 Rattus no
657	19	0.5	175103	9	HS37N07	AC353997 Homo sapi	730	19	0.5	186760	2	AC126996	AC126996 Rattus no
658	19	0.5	175112	9	AL138823	AL138823 Human DNA	731	19	0.5	187347	2	AC025880	AC025880 Homo sapi
659	19	0.5	175129	2	AC021490	AC021490 Homo sapi	c 732	19	0.5	187458	2	AC058863	AC058863 Homo sapi
660	19	0.5	175140	2	AC064823	AC064823 Homo sapi	733	19	0.5	187681	2	AC080124	AC080124 Homo sapi
661	19	0.5	176101	2	AF280813	AF280813 Homo sapi	734	19	0.5	187758	2	AC094868	AC094868 Rattus no
662	19	0.5	176261	8	AP003274	AP003274 Oryza sat	735	19	0.5	187767	2	AC123203	AC123203 Rattus no
663	19	0.5	176368	9	CNS01RGC	AL157971 Human chr	736	19	0.5	188246	2	AC013565	AC013565 Homo sapi
664	19	0.5	176601	2	AC026049	AC026049 Homo sapi	c 737	19	0.5	188431	9	AL162593	AL162593 Homo sapi
665	19	0.5	176723	2	AC128048	AC128048 Rattus no	c 738	19	0.5	188892	10	ALS91946	ALS91946 Mouse DNA
666	19	0.5	176889	9	AC090274	AC090274 Homo sapi	739	19	0.5	189250	9	AC074327	AC074327 Homo sapi
667	19	0.5	176935	8	AP001800	AP001800 Oryza sat	c 740	19	0.5	189403	2	AC128531	AC128531 Rattus no
668	19	0.5	177019	9	AC022833	AC022833 Homo sapi	741	19	0.5	189574	2	AL670100	AL670100 Mus muscu
669	19	0.5	177423	2	AC110297	AC110297 Homo sapi	c 742	19	0.5	189574	2	AC102574	AC102574 Mus muscu
670	19	0.5	177424	2	AC009832	AC009832 Homo sapi	c 743	19	0.5	189722	2	AC103275	AC103275 Rattus no
671	19	0.5	177544	2	AC124382	AC124382 Mus muscu	744	19	0.5	190033	2	AC019349	AC019349 Homo sapi
672	19	0.5	177552	2	AC097027	AC097027 Rattus no	745	19	0.5	190077	2	AC128021	AC128021 Rattus no
673	19	0.5	177568	9	HS388M5	Z97055 Human DNA s	746	19	0.5	191264	2	AC095203	AC095203 Rattus no
674	19	0.5	177703	9	AC019046	AC019046 Homo sapi	747	19	0.5	191343	2	AC108845	AC108845 Mus muscu
675	19	0.5	177891	2	AC117615	AC117615 Mus muscu	c 748	19	0.5	191676	2	AC105075	AC105075 Mus muscu
676	19	0.5	177893	9	AC005089	AC005089 Homo sapi	c 749	19	0.5	191911	9	AC098965	AC098965 Homo sapi
677	19	0.5	178078	2	AC127538	AC127538 Homo sapi	c 750	19	0.5	191918	2	AC026484	AC026484 Homo sapi
678	19	0.5	178236	2	AC112435	AC112435 Rattus no	751	19	0.5	192046	9	AC110772	AC110772 Homo sapi
679	19	0.5	178238	4	AC092519	AC092519 Felis cat	752	19	0.5	192087	2	AC079325	AC079325 Homo sapi
680	19	0.5	178364	2	AC021702	AC021702 Homo sapi	753	19	0.5	192337	9	AC022596	AC022596 Homo sapi
681	19	0.5	178621	2	AC021909	AC021909 Homo sapi	c 754	19	0.5	192356	2	AC102524	AC102524 Mus muscu
682	19	0.5	178669	2	AC016789	AC016789 Homo sapi	c 755	19	0.5	192518	2	AC113513	AC113513 Mus muscu
683	19	0.5	178918	2	AC131307	AC131307 Homo sapi	c 756	19	0.5	192725	9	AC092431	AC092431 Homo sapi
684	19	0.5	179009	9	AL355543	AL355543 Human DNA	c 757	19	0.5	193234	2	AC109535	AC109535 Rattus no
685	19	0.5	179098	2	AC117690	AC117690 Mus muscu	c 758	19	0.5	193487	2	AC127455	AC127455 Homo sapi
686	19	0.5	179141	2	AC019091	AC019091 Homo sapi	759	19	0.5	193852	9	AC099286	AC099286 Rattus no
687	19	0.5	179164	3	AC010692	AC010692 Drosophil	760	19	0.5	193893	9	AC008173	AC008173 Homo sapi
688	19	0.5	179372	2	AC040922	AC040922 Homo sapi	c 761	19	0.5	193897	2	AC069486	AC069486 Homo sapi
689	19	0.5	179463	2	AC094622	AC094622 Rattus no	c 762	19	0.5	194111	2	AC115190	AC115190 Rattus no
690	19	0.5	179569	10	AC118541	AC118541 Mus muscu	c 763	19	0.5	194133	2	AL824711	AL824711 Mus muscu
691	19	0.5	179611	2	AL833806	AL833806 Mus muscu	764	19	0.5	194762	2	AC094506	AC094506 Rattus no
692	19	0.5	179834	2	AC103659	AC103659 Mus muscu	c 765	19	0.5	194822	2	AC015818	AC015818 Homo sapi
693	19	0.5	180214	2	AC101751	AC101751 Mus muscu	766	19	0.5	195228	2	AC110779	AC110779 Homo sapi
694	19	0.5	180355	2	AC022167	AC022167 Homo sapi	c 767	19	0.5	195458	2	AC098258	AC098258 Rattus no
695	19	0.5	180465	9	AC074011	AC074011 Homo sapi	768	19	0.5	195673	2	AL807807	AL807807 Mus muscu
696	19	0.5	180517	2	AC099594	AC099594 Mus muscu	769	19	0.5	195779	2	AC094937	AC094937 Rattus no
697	19	0.5	180622	2	AC107835	AC107835 Mus muscu	770	19	0.5	195804	2	AL611952	AL611952 Mus muscu
698	19	0.5	180643	9	AC099555	AC099555 Papio cyn	771	19	0.5	195821	2	AC102085	AC102085 Mus muscu
699	19	0.5	180659	9	AC073648	AC073648 Homo sapi	c 772	19	0.5	195848	2	AC126513	AC126513 Rattus no
700	19	0.5	180790	9	AC012173	AC012173 Homo sapi	c 773	19	0.5	195964	2	AC114410	AC114410 Mus muscu
701	19	0.5	180855	2	AC128632	AC128632 Rattus no	c 774	19	0.5	196044	9	AC093844	AC093844 Homo sapi
702	19	0.5	180918	9	AC098680	AC098680 Homo sapi	c 775	19	0.5	196102	2	AC094425	AC094425 Rattus no
703	19	0.5	181073	2	AL844839	AL844839 Mus muscu	776	19	0.5	196216	2	AC019228	AC019228 Homo sapi
704	19	0.5	181243	10	AL645545	AL645545 Mouse DNA	c 777	19	0.5	196292	9	CNS0000B	AL049829 Human chr
705	19	0.5	181282	2	AC084240	AC084240 Rattus no	c 778	19	0.5	196333	2	AC094077	AC094077 Rattus no
706	19	0.5	181652	9	CNS01DUS	AL133367 Human chr	c 779	19	0.5	196349	2	AC019112	AC019112 Homo sapi
707	19	0.5	181853	2	AC111286	AC111286 Rattus no	c 780	19	0.5	196514	2	AC125555	AC125555 Rattus no
708	19	0.5	181912	2	AC095473	AC095473 Rattus no	c 781	19	0.5	196628	2	AL772388	AL772388 Danio rer
709	19	0.5	181998	2	AL805920	AL805920 Mus muscu	c 782	19	0.5	196641	9	AC093415	AC093415 Homo sapi
710	19	0.5	182228	2	AC115341	AC115341 Rattus no	783	19	0.5	196685	2	AC107555	AC107555 Rattus no
711	19	0.5	182591	2	AC097296	AC097296 Rattus no	c 784	19	0.5	197084	2	AC127663	AC127663 Rattus no
712	19	0.5	182666	2	AC129771	AC129771 Rattus no	785	19	0.5	197369	9	CNS01RH6	AL161662 Human chr
713	19	0.5	182807	2	AC026544	AC026544 Homo sapi	c 786	19	0.5	197498	9	AC005010	AC005010 Homo sapi
714	19	0.5	183672	9	CNS05TC8	AL355074 Human chr	787	19	0.5	198006	2	AL731852	AL731852 Mus muscu
715	19	0.5	183681	2	AC107814	AC107814 Mus muscu	c 788	19	0.5	198241	2	AP003721	AP003721 Homo sapi
716	19	0.5	183748	9	AC104964	AC104964 Homo sapi	c 789	19	0.5	198308	2	AC115542	AC115542 Rattus no
717	19	0.5	184140	9	AC090630	AC090630 Homo sapi	c 790	19	0.5	198838	2	AC093647	AC093647 Homo sapi
718	19	0.5	184305	2	AC125963	AC125963 Rattus no	c 791	19	0.5	199157	9	AC008555	AC008555 Homo sapi
719	19	0.5	184504	2	AL831716	AL831716 Mus muscu	792	19	0.5	199174	9	AC007742	AC007742 Homo sapi
720	19	0.5	184518	2	AC112120	AC112120 Rattus no	793	19	0.5	199474	10	AL672017	AL672017 Mouse DNA
721	19	0.5	184549	2	AC068575	AC068575 Homo sapi	794	19	0.5	199734	2	AC102141	AC102141 Mus muscu
722	19	0.5	184903	2	AC127445	AC127445 Rattus no	795	19	0.5	199790	2	AC128663	AC128663 Mus muscu

796	19	0.5	200000	9	AP000494	AP000494 Homo sapi	c 869	19	0.5	258761	2	AF181898	AF181898 Homo sapi
c 797	19	0.5	200062	4	AC091623	AC091623 Sus scrofa	c 870	19	0.5	259749	2	AL844852	AL844852 Mus muscu
798	19	0.5	200182	9	AC119037	AC119037 Homo sapi	c 871	19	0.5	263776	2	AC087335	AC087335 Mus muscu
c 799	19	0.5	200491	9	AC007249	AC007249 Homo sapi	c 872	19	0.5	264095	2	AC008500	AC008500 Homo sapi
800	19	0.5	200937	2	AC125667	AC125667 Rattus no	c 873	19	0.5	272258	2	AC123866	AC123866 Mus muscu
801	19	0.5	201106	2	AC110375	AC110375 Mus muscu	c 874	19	0.5	278572	10	AF287263	AF287263 Mus muscu
802	19	0.5	201642	2	AP005529	AP005529 Oryza sat	c 875	19	0.5	279965	3	AE003534	AE003534 Drosophila
803	19	0.5	201657	9	AL512427	AL512427 Human DNA	c 876	19	0.5	281662	2	AC117374	AC117374 Homo sapi
804	19	0.5	201707	2	AC131309	AC131309 Homo sapi	c 877	19	0.5	281892	2	AC125486	AC125486 Mus muscu
805	19	0.5	203200	9	AC008744	AC008744 Homo sapi	c 878	19	0.5	283568	2	AC099160	AC099160 Rattus no
c 806	19	0.5	203205	9	AC067750	AC067750 Homo sapi	c 879	19	0.5	300233	2	AC123842	AC123842 Mus muscu
c 807	19	0.5	204606	2	AC021131	AC021131 Homo sapi	c 880	19	0.5	302859	2	AC044785	AC044785 Homo sapi
c 808	19	0.5	204767	2	AC011367	AC011367 Homo sapi	c 881	19	0.5	312634	2	AC095096	AC095096 Rattus no
809	19	0.5	205557	2	AC102170	AC102170 Mus muscu	c 882	19	0.5	320679	2	AC126445	AC126445 Mus muscu
c 810	19	0.5	205736	2	AC124594	AC124594 Mus muscu	c 883	19	0.5	321944	2	AL353686	AL353686 Homo sapi
c 811	19	0.5	205835	2	AC126928	AC126928 Mus muscu	c 884	19	0.5	326150	1	AP005281	AP005281 Corynebact
c 812	19	0.5	206187	2	AC021172	AC021172 Homo sapi	c 885	19	0.5	335593	14	AF204951	AF204951 Ectocarpu
c 813	19	0.5	206373	2	AC122208	AC122208 Mus muscu	c 886	19	0.5	349980	6	AX127151	AX127151 Sequence
814	19	0.5	206456	2	AC112402	AC112402 Rattus no	c 887	19	0.5	349980	6	AX282123	AX282123 Sequence
815	19	0.5	206681	2	AL833794	AL833794 Mus muscu	c 888	19	0.5	349980	6	AX282124	AX282124 Sequence
816	19	0.5	206924	2	AL845498	AL845498 Mus muscu	c 889	19	0.5	349980	6	AX282715	AX282715 Sequence
817	19	0.5	207264	2	AC100750	AC100750 Mus muscu	c 890	19	0.5	349980	6	AX282716	AX282716 Sequence
818	19	0.5	207532	2	AC099298	AC099298 Rattus no	c 891	18	0.5	390	10	RATREPH	L24175 Rattus norv
819	19	0.5	208404	2	AC102234	AC102234 Mus muscu	c 892	18	0.5	199	11	HMSWX3105	L77626 Human chrom
c 820	19	0.5	208454	2	AC120128	AC120128 Mus muscu	c 893	18	0.5	235	5	AF227682	AF227682 Pseudupen
821	19	0.5	208571	9	AC010906	AC010906 Homo sapi	c 894	18	0.5	249	4	CEU11107	U11107 Cervus elap
822	19	0.5	208760	2	AC074048	AC074048 Mus muscu	c 895	18	0.5	249	4	CEU11108	U11108 Cervus elap
823	19	0.5	209162	2	AC090955	AC090955 Homo sapi	c 896	18	0.5	249	4	CEU11116	U11116 Cervus elap
c 824	19	0.5	209544	2	AC117629	AC117629 Mus muscu	c 897	18	0.5	249	4	CEU11117	U11117 Cervus elap
c 825	19	0.5	209636	2	AC121084	AC121084 Mus muscu	c 898	18	0.5	249	4	CEU11218	U11218 Cervus elap
c 826	19	0.5	209815	2	AL845290	AL845290 Mus muscu	c 899	18	0.5	249	4	CEU11230	U11230 Cervus elap
827	19	0.5	209816	2	AC105911	AC105911 Homo sapi	c 900	18	0.5	249	4	CEU11232	U11232 Cervus elap
c 828	19	0.5	210137	9	AC006504	AC006504 Homo sapi	c 901	18	0.5	273	5	AF369076	AF369076 Polyodon
c 829	19	0.5	210331	9	AL157402	AL157402 Human DNA	c 902	18	0.5	284	11	G30700	G30700 SWSS3415 Er
c 830	19	0.5	210500	2	AC109785	AC109785 Papio cyn	c 903	18	0.5	321	14	HS2PT034	HS2PT034
c 831	19	0.5	210790	2	AC126038	AC126038 Mus muscu	c 904	18	0.5	323	9	ABY13265	ABY13265 Acipenser b
c 832	19	0.5	212275	9	AC012076	AC012076 Homo sapi	c 905	18	0.5	342	5	AF395621	AF395621 Macaca mu
833	19	0.5	212347	2	AL772389	AL772389 Mus muscu	c 906	18	0.5	362	9	AF395622	AF395622 Macaca mu
c 834	19	0.5	212339	9	CNS01018B	AL121768 Human chr	c 907	18	0.5	362	9	AF395622	AF395622 Macaca mu
c 835	19	0.5	214738	30	AC023339	AC023339 Homo sapi	c 908	18	0.5	392	6	AX072105	AX072105 Sequence
c 836	19	0.5	215294	2	AC124362	AC124362 Mus muscu	c 909	18	0.5	397	9	F185592S19	F185592S19
837	19	0.5	215319	2	AC107798	AC107798 Mus muscu	c 910	18	0.5	406	14	BRU55253	BRU55253 Bovine resp
838	19	0.5	215563	2	AC125215	AC125215 Mus muscu	c 911	18	0.5	406	14	BRU55254	BRU55254 Bovine resp
839	19	0.5	217164	2	AC094182	AC094182 Rattus no	c 912	18	0.5	406	14	BRU55255	BRU55255 Bovine resp
840	19	0.5	218099	2	AC025504	AC025504 Homo sapi	c 913	18	0.5	406	14	BRU55256	BRU55256 Bovine resp
c 841	19	0.5	221372	2	AL772218	AL772218 Mus muscu	c 914	18	0.5	406	14	BRU55257	BRU55257 Bovine resp
c 842	19	0.5	223081	2	AC104908	AC104908 Mus muscu	c 915	18	0.5	418	3	AF379993	AF379993 Miropotes
c 843	19	0.5	223281	2	AC101671	AC101671 Mus muscu	c 916	18	0.5	418	3	AY044214	AY044214 Miropotes
844	19	0.5	223373	2	AC122881	AC122881 Mus muscu	c 917	18	0.5	419	9	AF395623	AF395623 Macaca mu
c 845	19	0.5	226962	2	AC079635	AC079635 Mus muscu	c 918	18	0.5	432	3	AF217339	AF217339 Archaster
c 846	19	0.5	226981	2	AC126550	AC126550 Mus muscu	c 919	18	0.5	439	6	AX261749	AX261749 Sequence
847	19	0.5	227208	2	AC096408	AC096408 Rattus no	c 920	18	0.5	450	11	G39092	G39092 213403 Zebr
848	19	0.5	227564	2	AC122804	AC122804 Mus muscu	c 921	18	0.5	450	14	AF248591	AF248591 Bovine re
c 849	19	0.5	228241	2	AC121906	AC121906 Mus muscu	c 922	18	0.5	450	14	AF248592	AF248592 Bovine re
850	19	0.5	228564	2	AC122251	AC122251 Mus muscu	c 923	18	0.5	450	14	AF248593	AF248593 Bovine re
851	19	0.5	230534	2	AC079575	AC079575 Mus muscu	c 924	18	0.5	450	14	AF248594	AF248594 Bovine re
852	19	0.5	230663	2	AC129219	AC129219 Mus muscu	c 925	18	0.5	450	14	AF248599	AF248599 Bovine re
c 853	19	0.5	234047	2	AC125061	AC125061 Mus muscu	c 926	18	0.5	450	14	AF248600	AF248600 Bovine re
c 854	19	0.5	236240	2	AL732590	AL732590 Mus muscu	c 927	18	0.5	450	14	AF248601	AF248601 Bovine re
c 855	19	0.5	236414	2	AC092742	AC092742 Mus muscu	c 928	18	0.5	450	14	AF248604	AF248604 Bovine re
856	19	0.5	236465	2	AL773565	AL773565 Mus muscu	c 929	18	0.5	450	14	AF248607	AF248607 Bovine re
c 857	19	0.5	237405	2	AC068142	AC068142 Mus muscu	c 930	18	0.5	453	14	AF248605	AF248605 Bovine re
c 858	19	0.5	239847	10	AL591466	AL591466 Mouse DNA	c 931	18	0.5	461	11	G33671	G33671 human STS W
859	19	0.5	242617	2	AL844570	AL844570 Danio rer	c 932	18	0.5	492	14	AF188577	AF188577 Bovine re
c 860	19	0.5	243290	10	AL663088	AL663088 Mouse DNA	c 933	18	0.5	492	14	AF188578	AF188578 Bovine re
c 861	19	0.5	243979	2	AC097077	AC097077 Rattus no	c 934	18	0.5	492	14	AF188579	AF188579 Bovine re
862	19	0.5	245844	2	AC021094	AC021094 Homo sapi	c 935	18	0.5	492	14	AF188580	AF188580 Bovine re
c 863	19	0.5	247226	2	AC122936	AC122936 Mus muscu	c 936	18	0.5	492	14	AF188581	AF188581 Bovine re
864	19	0.5	248361	2	AC124579	AC124579 Mus muscu	c 937	18	0.5	492	14	AF188582	AF188582 Bovine re
c 865	19	0.5	248862	2	AC091215	AC091215 Rattus no	c 938	18	0.5	492	14	AF188583	AF188583 Bovine re
c 866	19	0.5	248974	2	AC103324	AC103324 Rattus no	c 939	18	0.5	492	14	AF188584	AF188584 Bovine re
867	19	0.5	255370	2	AC110031	AC110031 Mus muscu	c 940	18	0.5	494	8	LHA410056	AJ410056 Leptadeni
c 868	19	0.5	256657	2	AC129177	AC129177 Mus muscu	c 941	18	0.5	495	8	AF077632	AF077632 Couroupit

QY 61 TTAGCTGAACACAGACATCGGGAAACAATGAGAGAGTCGAGCGGAGCCCTCCTCCAA 120
|||||
Db 548 TTAGCTGAACACAGACATCGGGAAACAATGAGAGAGTCGAGCGGAGCCCTCCTCCAA 607
|||||
QY 121 CCTTGGCAATTTGGCAGTCTCTTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 180
|||||
Db 608 CCTTGGCAATTTGGCAGTCTCTTGGGGCCGCTGAAGGAGCGGAGGCCAAGATGACCTT 667
|||||
QY 181 CCAGATCTTTTCAGCTCTTCTGAGCCAAAGAAATAGACGAAGTGTCAATTTGGCAAGA 240
|||||
Db 668 CCAGATCTTTTCAGCTCTTCTGAGCCAAAGAAATAGACGAAGTGTCAATTTGGCAAGA 727
|||||
QY 241 CTGGCCATCAATTACGACCTTTTGGAGAAGCGCAGATGAACCTCAAGCTAGAAAACGACTT 300
|||||
Db 728 CTGGCCATCAATTACGACCTTTTGGAGAAGCGCAGATGAACCTCAAGCTAGAAAACGACTT 787
|||||
QY 301 TCTCCTGATCAGATGAACACACTCACCTAATTTAAAGTTTGGAGCCTAACCTTCTGCCAGGAT 360
|||||
Db 788 TCTCCTGATCAGATGAACACACTCACCTAATTTAAAGTTTGGAGCCTAACCTTCTGCCAGGAT 847
|||||
QY 361 AACCCCTCGAAGTCCCAACAGCTCTAAAGAAAGCCCGCAGGAGGCAAAAAGGCCACAGTAT 420
|||||
Db 848 AACCCCTCGAAGTCCCAACAGCTCTAAAGAAAGCCCGCAGGAGGCAAAAAGGCCACAGTAT 907
|||||
QY 421 TGTCTGAAACCCAGTCCAAAAGATTTTAAATTAAGGCTGCCGACTTCATTGGAAGAG 480
|||||
Db 908 TGTCTGAAACCCAGTCCAAAAGATTTTAAATTAAGGCTGCCGACTTCATTGGAAGAG 967
|||||
QY 481 CTATCCTCCCTTTTCAATTCACAGCTCCAAAGATTTAGACCTCGTCCGCTGCAAAAAC 540
|||||
Db 968 CTATCCTCCCTTTTCAATTCACAGCTCCAAAGATTTAGACCTCGTCCGCTGCAAAAAC 1027
|||||
QY 541 CACAAGAGTAACTGGAATCTCAAAAACAAGTTATGCGAGAAACAGCTCCAGTTTCTCA 600
|||||
Db 1028 CACAAGAGTAACTGGAATCTCAAAAACAAGTTATGCGAGAAACAGCTCCAGTTTCTCA 1087
|||||
QY 601 GATCTGCAAGAAAGCAGAGAAGATCTCTGTGTCCCATCCCTATCCCTCGGGATACCAGG 660
|||||
Db 1088 GATCTGCAAGAAAGCAGAGAAGATCTCTGTGTCCCATCCCTATCCCTCGGGATACCAGG 1147
|||||
QY 661 GATAATGAAGTGAATCAGCCCTGGAAACAGCAGGAAGCCAGAGGCGTGAAGCGGAGCAG 720
|||||
Db 1148 GATAATGAAGTGAATCAGCCCTGGAAACAGCAGGAAGCCAGAGGCGTGAAGCGGAGCAG 1207
|||||
QY 721 GCTGCCAGTGAAGCGGCTGTGTGAGACACTACACAGGCTTTCCTCCCTTCATCTCTGTAC 780
|||||
Db 1208 GCTGCCAGTGAAGCGGCTGTGTGAGACACTACACAGGCTTTCCTCCCTTCATCTCTGTAC 1267
|||||
QY 781 TATGAAGAACCTCTGCGGCAACCTCCCGGTTTCACCTCAAAAAGTTACGGAGCAGAGAAGTT 840
|||||
Db 1268 TATGAAGAACCTCTGCGGCAACCTCCCGGTTTCACCTCAAAAAGTTACGGAGCAGAGAAGTT 1327
|||||
QY 841 CCAGAAGGAACCTCGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACACCTCAAGTA 900
|||||
Db 1328 CCAGAAGGAACCTCGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACACCTCAAGTA 1387
|||||
QY 901 AGGTGTACTGTGAAGCAGAGGCTTGAATTTCCCAAGATATTCACATCGTCCAGGCA 960
|||||
Db 1388 AGGTGTACTGTGAAGCAGAGGCTTGAATTTCCCAAGATATTCACATCGTCCAGGCA 1447
|||||
QY 961 GGAATCTGCATCTACTGACCATTTGCGGAAGCCTTTGAAGAGGACACAGSACGCTATTCC 1020
|||||
Db 1448 GGAATCTGCATCTACTGACCATTTGCGGAAGCCTTTGAAGAGGACACAGSACGCTATTCC 1507
|||||
QY 1021 TGCCTTGTCTTAACATCTATGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1080
|||||
Db 1508 TGCCTTGTCTTAACATCTATGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1567
|||||
QY 1081 GTTCTCTTCTCTGACTCAGAGGCGACCTTAACAAGGAGAGATGAATCAATCCAGAAG 1140
|||||
Db 1568 GTTCTCTTCTCTGACTCAGAGGCGACCTTAACAAGGAGAGATGAATCAATCCAGAAG 1627
|||||
QY 1141 CCAAAATGAGGTGTCACTCTCTCCCACTACCTCTGACAGTCAATTCCTCCAGCAGTACCCCAA 1200
|||||

Db 1628 CCAAAATGAGGTGTCACTCTCTCCCACTACCTCTGCACTCATTCCTCCAGCAGTACCCCAA 1687
|||||
QY 1201 GCCCAGCATTTTGGTGCCCAAACTCGTGTGGCAACCATCCAGCAGTGTGAGAGCCCAACC 1260
|||||
Db 1688 GCCCAGCATTTTGGTGCCCAAACTCGTGTGGCAACCATCCAGCAGTGTGAGAGCCCAACC 1747
|||||
QY 1261 AATTACTTCCAGGGATTGGATGGAACACCTATCATTCGACGCTCCTGTGTTCACAAAGATG 1320
|||||
Db 1748 AATTACTTCCAGGGATTGGATGGAACACCTATCATTCGACGCTCCTGTGTTCACAAAGATG 1807
|||||
QY 1321 CTACAAAATTTGTACGTTCTGAGGTCAGTGGTTGTCTTTGAATGCAAGTAAAGA 1380
|||||
Db 1808 CTACAAAATTTGTACGTTCTGAGGTCAGTGGTTGTCTTTGAATGCAAGTAAAGA 1867
|||||
QY 1381 GCTCCATCTCTTAAGSTTGTAGTGGTATAGAGAAGGACCTTTAATAGAATCTTCAGAT 1440
|||||
Db 1868 GCTCCATCTCTTAAGSTTGTAGTGGTATAGAGAAGGACCTTTAATAGAATCTTCAGAT 1927
|||||
QY 1441 TTTAGGATTTTACAGAAAACCTCGATCCATGGCAGAGCCAGAGGAGATTTTCACCTTG 1500
|||||
Db 1928 TTTAGGATTTTACAGAAAACCTCGATCCATGGCAGAGCCAGAGGAGATTTTCACCTTG 1987
|||||
QY 1501 GTCAATTTGCTGAGGTGTTTGCAGAAAGATTTCTGGGTGCTTCACATGTACTGCAAGCAACAAA 1560
|||||
Db 1988 GTCAATTTGCTGAGGTGTTTGCAGAAAGATTTCTGGGTGCTTCACATGTACTGCAAGCAACAAA 2047
|||||
QY 1561 TAGGGCAGAGTGTCAAGCATTTGCACAGCTGCACAGCTGAGAGGAAATGAGACCTTCAGCAAC 1620
|||||
Db 2048 TAGGGCAGAGTGTCAAGCATTTGCACAGCTGCACAGCTGAGAGGAAATGAGACCTTCAGCAAC 2107
|||||
QY 1621 AAGGGTCTCTTCACATCAGCCTCACTACCACTGACAGCTATTTGAGGCCACAGCCC 1680
|||||
Db 2108 AAGGGTCTCTTCACATCAGCCTCACTACCACTGACAGCTATTTGAGGCCACAGCCC 2167
|||||
QY 1681 TCCCCACCCCACTCAGAGCTTCCATCTGTGTGAACAAACCCCAACCCCAACCCCAACCCCAAGGGG 1740
|||||
Db 2168 TCCCCACCCCACTCAGAGCTTCCATCTGTGTGAACAAACCCCAACCCCAACCCCAACCCCAAGGGG 2227
|||||
QY 1741 GTTCTGCTGAACACCAATGAGCCCGCTCAGCTCCAGATTTGGGCTTCGTGTGCACTTC 1800
|||||
Db 2228 GTTCTGCTGAACACCAATGAGCCCGCTCAGCTCCAGATTTGGGCTTCGTGTGCACTTC 2287
|||||
QY 1801 AACCTGCCTGAAGTACAAAGGAAGTGAAGCATCTCCGAGGCTGTGTGTGTGACCAACC 1860
|||||
Db 2288 AACCTGCCTGAAGTACAAAGGAAGTGAAGCATCTCCGAGGCTGTGTGTGTGACCAACC 2347
|||||
QY 1861 AGACAGACAGGCGCCGATTTTSCAGGAGAGTTCAAGGACAGGCAACAAACCAACCAACCA 1920
|||||
Db 2348 AGACAGACAGGCGCCGATTTTSCAGGAGAGTTCAAGGACAGGCAACAAACCAACCAACCA 2407
|||||
QY 1921 GAGCCTTCTTCCCGTGAAGAGCCCTCCAGTTCTGGCCAAACCCCAACCTTGATTTC 1980
|||||
Db 2408 GAGCCTTCTTCCCGTGAAGAGCCCTCCAGTTCTGGCCAAACCCCAACCTTGATTTC 2467
|||||
QY 1981 ACTCAGTTCAACAGCTTCAATAACCAAGTCTTACTGGAACAACCAACCAATTCGCAAAACCCA 2040
|||||
Db 2468 ACTCAGTTCAACAGCTTCAATAACCAAGTCTTACTGGAACAACCAACCAATTCGCAAAACCCA 2527
|||||
QY 2041 CTTCTCTTCACTCTTAAGGAGTTTCTTCARCATGACTTTTGAACCTCAATGCTTCC 2100
|||||
Db 2528 CTTCTCTTCACTCTTAAGGAGTTTCTTCARCATGACTTTTGAACCTCAATGCTTCC 2587
|||||
QY 2101 CCAGCGGTGACACATCCCAATAGCAGGTGAAGCTCTTCATCAGAGAGGTTTCAGCTTG 2160
|||||
Db 2588 CCAGCGGTGACACATCCCAATAGCAGGTGAAGCTCTTCATCAGAGAGGTTTCAGCTTG 2647
|||||
QY 2161 GCCCGCGCAAGTATTTCTCCCTCCACGAACACACCCAGCAACTGTGGCCCTTCC 2220
|||||
Db 2648 GCCCGCGCAAGTATTTCTCCCTCCACGAACACACCCAGCAACTGTGGCCCTTCC 2707
|||||
QY 2221 AGCTCTCCGGTGTCACTTTGAGCAGCACTCTCAAACTATTTCAGAGGACAGTGAACAAA 2280
|||||

Db	2708	AGCTCTCGGTGTTTCACTTTTGAGCAGCACTCTCTCAAACTATTTCAGAGGACAGTTCAGCAAA	27676
Qy	2281	GAAGCCTCTTTAGTGTCTCACCCCTCTGTGCAAAACAAATCTCCAGGAGGCTTTTCCATC	2340
Db	2768	GAAGCCTCTTTAGTGTCTCACCCCTCTGTGCAAAACAAATCTCCAGGAGGCTTTTCCATC	2827
Qy	2341	CAAAATGAGCCACTCCACCAGGCCCAACAGAACCRACACCAACCAATTCACATTTTCC	2400
Db	2828	CAAAATGAGCCACTCCACCAGGCCCAACAGAACCAACCAACCAACCAATTCACATTTTCC	2887
Qy	2401	ATCCCCAGCGAAACCAAGTTTCAGCGCCGCTGTGTGCCCAATTCCTGTCTCTCTTACC	2460
Db	2888	ATCCCCAGGGAAACCAAGTTTCAGCGCCGCTGTGTGCCCAATTCCTGTCTCTCTTACC	2947
Qy	2461	AGCCGGATTCAGAACCCAGTGGCTTTCACGCTCTGTGTGCTTGCCTCTCTCCCTGCCATC	2520
Db	2948	AGCCGGATTCAGAACCCAGTGGCTTTCACGCTCTGTGTGCTTGCCTCTCTCCCTGCCATC	3007
Qy	2521	CCACCACAAATGCCATGGRGCTGCTTGAAGTGCACCATCATGTGCCATCCACAGGATTA	2580
Db	3008	CCACCACAAATGCCATGGGCTGCTTGAAGTGCACCATCATGTGCCATCCACAGGATTA	3067
Qy	2581	GCGAAGAAAAATACAAGTCTCCTCAACCAAGTGAATGATGATAAATTCGTGAAACTAAG	2640
Db	3068	GCGAAGAAAAATACAAGTCTCCTCAACCAAGTGAATGATGATAAATTCGTGAAACTAAG	3127
Qy	2641	AACGCACTGATTCGAGACTTGGGGAAAAAATAACTTTTCAGTGATGTCCAGCAACCAACG	2700
Db	3128	AACGCACTGATTCGAGACTTGGGGAAAAAATAACTTTTCAGTGATGTCCAGCAACCAACG	3187
Qy	2701	CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAATAGAGATTTTCGCTTG	2760
Db	3188	CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAATAGAGATTTTCGCTTG	3247
Qy	2761	GAAGCTACTCCTGTTGATGAATCAGATGATGAATTCACATGATGAGATCCCCACGGCC	2820
Db	3248	GAAGCTACTCCTGTTGATGAATCAGATGATGAATTCACATGATGAGATCCCCACGGCC	3307
Qy	2821	AAGTGTATTGCTCCCATCTTTTACAAAGACATCAAGCACTTTCCGGGTACAGAAAGGCTCT	2880
Db	3308	AAGTGTATTGCTCCCATCTTTTACAAAGACATCAAGCACTTTCCGGGTACAGAAAGGCTCT	3367
Qy	2881	CCAGTCAATTCACCTGCAAAATTTGTTGGGTACCTGTTTCAAGGTTTACTGGTTCAAA	2940
Db	3368	CCAGTCAATTCACCTGCAAAATTTGTTGGGTACCTGTTTCAAGGTTTACTGGTTCAAA	3427
Qy	2941	GATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGCAAAATGAGCGAGAAGAGATGGG	3000
Db	3428	GATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGCAAAATGAGCGAGAAGAGATGGG	3487
Qy	3001	ACATGCTCTCTGCACATTTGAATCCACTACCATGATGACGATGGCAACTACACCATCATG	3060
Db	3488	ACATGCTCTCTGCACATTTGAATCCACTACCATGATGACGATGGCAACTACACCATCATG	3547
Qy	3061	GCAGCCAAACCCACAGGGAGATCAGCTGCTCTGCGCACTTGATGGTACAAGTTTGGCC	3120
Db	3548	GCAGCCAAACCCACAGGGAGATCAGCTGCTCTGCGCACTTGATGGTACAAGTTTGGCC	3607
Qy	3121	ATTGCGAGTCGGGTAACTCTCTGCTGCTAGTCTACACAGGGGAAGATCCCGAGTCAAGAA	3180
Db	3608	ATTGCGAGTCGGGTAACTCTCTGCTGCTAGTCTACACAGGGGAAGATCCCGAGTCAAGAA	3667
Qy	3181	AGAGCAAAAGAGCCCTACAGNAACGCTTTTTCGACCAACATTTCTCTCGAGGCTCTGGG	3240
Db	3668	AGAGCAAAAGAGCCCTACAGNAACGCTTTTTCGACCAACATTTCTCTCGAGGCTCTGGG	3727
Qy	3241	GATATGTTAGCTCATGAGGGGGCCCTCTCTGCTGGCTGGACTGTAAAGGTGAGTGGTTACCG	3300
Db	3728	GATATGTTAGCTCATGAGGGGGCCCTCTCTGCTGGCTGGACTGTAAAGGTGAGTGGTTACCG	3787
Qy	3301	CCCCGGAGCTGACATGGCTACTCAATAGCCCAACCTGTCTACCAGATGCTCTCCCAAG	3360
Db	3788	CCCCGGAGCTGACATGGGTACTCAATGGCCAACTGTCTACAGATGCTCTCCCAAG	3847

QY	3361	ATGCTGGTCAAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACACTACTCAGCGGCAC	3420
DB	3848	ATGCTGGTCAAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACACTACTCAGCGGCAC	3907
QY	3421	GCAGGGACCTTAAAGTCATCGCTACCAACAAAACCGGGCAGAAATCTTTTAGTCTCGAG	3480
DB	3908	GCAGGGACCTTAAAGTCATCGCTACCAACAAAACCGGGCAGAAATCTTTTAGTCTCGAG	3967
QY	3481	CTCTCTGTAGTAGCCAAAGAGGTGAAGAAAGACACTGTGATCCTCGGAAACTACAGAAC	3540
DB	3968	CTCTCTGTAGTAGCCAAAGAGGTGAAGAAAGACACTGTGATCCTCGGAAACTACAGAAC	4027
QY	3541	TGCGGTGTTCCGGAAGGCCACCCGCTGAGACTGAGAGTGCCTGCTCATTCAGCCAGCCAGAAATCAGAC	3600
DB	4028	TGCGGTGTTCCGGAAGGCCACCCGCTGAGACTGAGAGTGCCTGCTCATTCAGCCAGCCAGAAATCAGAC	4087
QY	3601	CCTGTGTTCTACTGGAAGAAAGACAAATGAGACCATCCCTTGCACACAGAGAGATCAGT	3660
DB	4088	CCTGTGTTCTACTGGAAGAAAGACAAATGAGACCATCCCTTGCACACAGAGAGATCAGT	4147
QY	3661	ATGCACACAGGACACAAACAGGTATGCCTGCTCTCTCATTCAGCCAGCCAGAAATCAGAC	3720
DB	4148	ATGCACACAGGACACAAACAGGTATGCCTGCTCTCTCATTCAGCCAGCCAGAAATCAGAC	4207
QY	3721	GCTGATGGTACAGTTGTCAGCAAGAAATGAAGCCGGCATCGTGTGCATGTCACAG	3780
DB	4208	GCTGATGGTACAGTTGTCAGCAAGAAATGAAGCCGGCATCGTGTGCATGTCACAG	4267
QY	3781	CTGGATATATACGCTACGTGGCACCACATCAGATCCACCGCCCATGCTGTGTCGGGCCAGT	3840
DB	4268	CTGGATATATACGCTACGTGGCACCACATCAGATCCACCGCCCATGCTGTGTCGGGCCAGT	4327
QY	3841	GGCAGTCTCGTACGGATCTCTCACCAGTAAAGGACTTGACATATTTTCTGCGCTTTTCCTCC	3900
DB	4328	GGCAGTCTCGTACGGATCTCTCACCAGTAAAGGACTTGACATATTTTCTGCGCTTTTCCTCC	4387
QY	3901	ATGGAACACAGATGGTGTATTCATGCTCTCTCGGAGTGTAGTGGAGAGTGATGAACCTT	3960
DB	4388	ATGGAACACAGATGGTGTATTCATGCTCTCTCGGAGTGTAGTGGAGAGTGATGAACCTT	4447
QY	3961	TAA 3963	
DB	4448	TAA 4450	
RESULT 2			
LOCUS	HS805275	5804 bp mRNA linear	PRI 12-JUL-2002
DEFINITION	Homo sapiens mRNA; cDNA DKF2p451E012 (from clone DKF2p451E012).		
ACCESSION	AL834247		
VERSION	AL834247.1	GI:21739789	
KEYWORDS	human.		
SOURCE			
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 5804)		
JOURNAL	Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY		
COMMENT	Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by AGOFA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKF2p451E012) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/ . Location/Qualifiers		
FEATURES			
source	1. 5804		
	/organism="Homo sapiens"		

[illegible]

Db	2725	AGCCGGATT	CAGAACCAAGTGGCTTTCCTCAGCTCTGTGTGCGCTTCTCTCCCTCGCCATC	2780
Qy	2521	CAACCCAAATGCCATGGRGCTGCCCTAGAAGTGCACCATCCATGCCATCCACAGGATTA	2580	
Db	2785	CAACCCAAATGCCATGGGGCTGCGCTACAAGTGCACCATCCATGCCATCCACAGGATTA	2844	
Qy	2581	CGAAGAAAAATACAAGTCTCCTCAACCAAGTGAATGATGATACATTCGTGAAACTAAG	2640	
Db	2845	CGAAGAAAAATACAAGTCTCCTCAACCAAGTGAATGATGATACATTCGTGAAACTAAG	2904	
Qy	2641	AACCCAGTGATTCGAGACTTGGGGAAAAAATAACTTTCAGTGTAGTGCAGACCAAAACAG	2700	
Db	2905	AACCCAGTGATTCGAGACTTGGGGAAAAAATAACTTTCAGTGTAGTGCAGACCAAAACAG	2964	
Qy	2701	CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAATAGAGTTTCGCTTG	2760	
Db	2965	CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAATAGAGTTTCGCTTG	3024	
Qy	2761	GAACGTACTCTGTGTGATGAATCAGATGATGAATAATTCAACATGATGAGATCCCAACGGC	2820	
Db	3025	GAACGTACTCTGTGTGATGAATCAGATGATGAATAATTCAACATGATGAGATCCCAACGGC	3084	
Qy	2821	AAGTGTATTGCTCCCATCTTTGACAAGAGACTCAAGCACTTCCGGGTCACAGAAGGCTCT	2880	
Db	3085	AAGTGTATTGCTCCCATCTTTGACAAGAGACTCAAGCACTTCCGGGTCACAGAAGGCTCT	3144	
Qy	2881	CCAGTCACTTTCACCTGCAAAATTTGTTGGGATACCTGTTCCAAAGTTTACTGGTTCAAA	2940	
Db	3145	CCAGTCACTTTCACCTGCAAAATTTGTTGGGATACCTGTTCCAAAGTTTACTGGTTCAAA	3204	
Qy	2941	GATGGGAAGCAGATTTCCTAAGGAAATGAGCACTGC AAAATGAGCGCAGAGAGATGGG	3000	
Db	3205	GATGGGAAGCAGATTTCCTAAGGAAATGAGCACTGC AAAATGAGCGCAGAGAGATGGG	3264	
Qy	3001	ACATGCTCTCTGCACATTTGAATCCACTACCATGATGACGATGGCAACTACACCATCATG	3060	
Db	3265	ACATGCTCTCTGCACATTTGAATCCACTACCATGATGACGATGGCAACTACACCATCATG	3324	
Qy	3061	GCAGCAACCCCAAGGGGAGAAATCAGCTCTCTGCGCCACTTTGATGTTACAAAGTTTCCGC	3120	
Db	3325	GCAGCAACCCCAAGGGGAGAAATCAGCTCTCTGCGCCACTTTGATGTTACAAAGTTTCCGC	3384	
Qy	3121	ATTCGCAGTCGGCTAACCTCTCTGTGTGCTCAGTCTCACAGGGGAAGATCCCGAGTCAAGAA	3180	
Db	3385	ATTCGCAGTCGGCTAACCTCTCTGTGTGCTCAGTCTCACAGGGGAAGATCCCGAGTCAAGAA	3444	
Qy	3181	AGACAAAGAGCCCTTACAGGAACGCTTTTTCGACACACATTTCTCTGCAGGCTCTCTGGG	3240	
Db	3445	AGACAAAGAGCCCTTACAGGAACGCTTTTTCGACACACATTTCTCTGCAGGCTCTCTGGG	3504	
Qy	3241	GATATGTAGCTCATGAGGGGCCCTCTCTGCGCTGGACTGTAAAGTGAGTGGTTTACCG	3300	
Db	3505	GATATGTAGCTCATGAGGGGCCCTCTCTGCGCTGGACTGTAAAGTGAGTGGTTTACCG	3564	
Qy	3301	CCCCGGAGGTGACATGGCTACTCAATGGCCAAACCTGTGCTACCAGATGCCTCCCAACAG	3360	
Db	3565	CCCCGGAGGTGACATGGCTACTCAATGGCCAAACCTGTGCTACCAGATGCCTCCCAACAG	3624	
Qy	3361	ATGCTGTGACAGGAGACCGGAGTCCACTCTCTGTCTCATTTGACCACACTCATCAGCGGCAC	3420	
Db	3625	ATGCTGTGACAGGAGACCGGAGTCCACTCTCTGTCTCATTTGACCACACTCATCAGCGGCAC	3684	
Qy	3421	GCAGGGACCTTAAAGTGATCAGCTACCAACAAAAACCGGGCAGAAATCTTTTAGTCTGGAG	3480	
Db	3685	GCAGGGACCTTAAAGTGATCAGCTACCAACAAAAACCGGGCAGAAATCTTTTAGTCTGGAG	3744	
Qy	3481	CTCTCTCTAGTACGCAAAAGAGTGAAGAAGCAACCTGTGATCCTGGAGAAACTACAGAAC	3540	
Db	3745	CTCTCTCTAGTACGCAAAAGAGTGAAGAAGCAACCTGTGATCCTGGAGAAACTACAGAAC	3804	
Qy	3541	TGCGGTGTTCCGAAGGCCACCCCGTGAGACTGGAGTGCCCGGTGATAGGCATGCCCCCA	3600	
Db	3805	TGCGGTGTTCCGAAGGCCACCCCGTGAGACTGGAGTGCCCGGTGATAGGCATGCCCCCA	3864	

```
QY 3601 CCTGTGTTCTACTGGAGAAAGACAATGAGACCATCCCTTGCACACGAGAGAGGATCAGT 3660
|||||
Db 3865 CCTGTGTTCTACTGGAGAAAGACAATGAGACCATCCCTTGCACACGAGAGAGATCAGT 3924
|||||
QY 3661 ATGCACAGAGACACAACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATCAGAC 3720
|||||
Db 3925 ATGCACAGAGACACAACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATCAGAC 3984
|||||
QY 3721 GCTGGATGTACACGTTGTTCAGCCCAAGAAATGAAGCCGGCATCTGTCTGTCACCTGCCAGG 3780
|||||
Db 3985 GCTGGATGTACACGTTGTTCAGCCCAAGAAATGAAGCCGGCATCTGTCTGTCACCTGCCAGG 4044
|||||
QY 3781 CTGGATATATACGCTCAGTGGCCACCATCAGATCCCAACGCCCATGTCTGTCCGGCCCATG 3840
|||||
Db 4045 CTGGATATATACGCTCAGTGGCCACCATCAGATCCCAACGCCCATGTCTGTCCGGCCCATG 4104
|||||
QY 3841 GGCAGTCGCTACGGATCTCTCAGCAGTAAGACTTGACATATTTCTGCTCTTTCTCTCC 3900
|||||
Db 4105 GGCAGTCGCTACGGATCTCTCAGCAGTAAGACTTGACATATTTCTGCTCTTTCTCTCC 4164
|||||
QY 3901 ATGGAAGCAGCATGTGTATTCATCTCTCTCGAGTGTCTAGTGAGAGTGATGAACATT 3960
|||||
Db 4165 ATGGAAGCAGCATGTGTATTCATCTCTCTCGAGTGTCTAGTGAGAGTGATGAACATT 4224
|||||
QY 3961 TAA 3963
|||
Db 4225 TAA 4227
|||

RESULT 3
HSM803309 5486 bp mRNA linear PRI 12-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp451M023 (from clone DKFZp451M023).
DEFINITION AL832002
ACCESSION
VERSION AL832002.1 GI:21732542
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5486)
Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.
Direct Submission
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp451M023) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cbna/.
FEATURES
Location/Qualifiers
source
1..5486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp451M023"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
pH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal 5451..5456
polyA_site 5468
BASE COUNT 1558 a 1392 c 1260 g 1276 t
ORIGIN
Query Match 82.0%; Score 3249; DB 9; Length 5486;
Best Local Similarity 99.6%; Pred.No. 0;
Matches 3949; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 ATGCAAGACGACAGCATAGAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 60
|||||
```

```
Db 266 ATGCAAGACGACAGCATAGAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 325
|||||
QY 61 TTAGCTGAARACACAGACATCGGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCTCAAC 120
|||||
Db 326 TTAGCTGAARACACAGACATCGGGGAAACAATGAGAGAGTCGAGCGAGCCCTCTCTCAAC 385
|||||
QY 121 CCTTGGCATTTCCGGCAGTCTCTTGGGGCGCTGAAGGAGCGGAGCCCAAGATGACCTT 180
|||||
Db 386 CCTTGGCATTTCCGGCAGTCTCTTGGGGCGCTGAAGGAGCGGAGCCCAAGATGACCTT 445
|||||
QY 181 CCAGATCTTTTACGCCCTTTCTGAGCCAAGAAGAATTAGACGAAAGTGTCTCAATTTGGCAAGA 240
|||||
Db 446 CCAGATCTTTTACGCCCTTTCTGAGCCAAGAAGAATTAGACGAAAGTGTCTCAATTTGGCAAGA 505
|||||
QY 241 CTGGCCATCAATTTACGACCCCTTTTGGAGAGGCGAGATCAAACTCAAGCTAGAAAACGACTT 300
|||||
Db 506 CTGGCCATCAATTTACGACCCCTTTTGGAGAGGCGAGATGAAACTCAAGCTAGAAAACGACTT 565
|||||
QY 301 TCTCCTGATCAGATGAAACACTCACCTTAATTTAAGTTTTTTCAGGCTTAACCTTCTGCCAGGAT 360
|||||
Db 566 TCTCCTGATCAGATGAAACACTCACCTTAATTTAAGTTTTTTCAGGCTTAACCTTCTGCCAGGAT 625
|||||
QY 361 AACCCCTCGAAGTCCACACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAAGGCCACAGTAT 420
|||||
Db 626 AACCCCTCGAAGTCCACACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAAGGCCACAGTAT 685
|||||
QY 421 TGTTCCTGAACCCAGCTCCAAAAGATTAATTTAAATAAGGCTGCCGACTTCATTTGAAGAG 480
|||||
Db 686 TGTTCCTGAACCCAGCTCCAAAAGATTAATTTAAATAAGGCTGCCGACTTCATTTGAAGAG 745
|||||
QY 481 CTATCCTCCCTTTTCAAAATCCCAAGCTCCAAAAGATTAAGACTCGTCTGCTGCAAAAAC 540
|||||
Db 746 CTATCCTCCCTTTTCAAAATCCCAAGCTCCAAAAGATTAAGACTCGTCTGCTGCAAAAAC 805
|||||
QY 541 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTATGCAAGGAAACACAGCTCCAGTTTCTCA 600
|||||
Db 806 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTATGCAAGGAAACACAGCTCCAGTTTCTCA 865
|||||
QY 601 GATCTGTCAGAAAGACGAGAAAGATCTTCTGTTCCCATCCCTATCCCTGCGGATACACAGG 660
|||||
Db 866 GATCTGTCAGAAAGACGAGAAAGATCTTCTGTTCCCATCCCTATCCCTGCGGATACACAGG 925
|||||
QY 661 GATAATGAAGTGAATCACGCGCTGGAAACACGAGCAAGAGCGCTGAAGCGGAGCAG 720
|||||
Db 926 GATAATGAAGTGAATCACGCGCTGGAAACACGAGCAAGAGCGCTGAAGCGGAGCAG 985
|||||
QY 721 GCTGCCAGTCAGCGCGCTGGTGAGACACTACACAGGGTCTTTCCCTCTCATCTCTGTAC 780
|||||
Db 986 GCTGCCAGTCAGCGCGCTGGTGAGACACTACACAGGGTCTTTCCCTCTCATCTCTGTAC 1045
|||||
QY 781 TATGAAGAACCTCTGGGCAACCTCCCGGTTCACTCAAAAGTTACGAGCAGAGAAGTT 840
|||||
Db 1046 TATGAAGAACCTCTGGGCAACCTCCCGGTTCACTCAAAAGTTACGAGCAGAGAAGTT 1105
|||||
QY 841 CCAGAGGAACCTCGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACACACCTCAAGTA 900
|||||
Db 1106 CCAGAGGAACCTCGAGTACAGTTGGATTGCATAGTGGTAGGAATTCACACACCTCAAGTA 1165
|||||
QY 901 AGTGGTACTGTGAAGCAAGAGACTTGAATAATTTCCCGAGATATTCACATCGTCCAGGCA 960
|||||
Db 1166 AGTGGTACTGTGAAGCAAGAGACTTGAATAATTTCCCGAGATATTCACATCGTCCAGGCA 1225
|||||
QY 961 GGAATCTGCACCTCACTGACCATTCGCGAAGCCCTTGAAGAGGACACAGACGCTATTC 1020
|||||
Db 1226 GGAATCTGCACCTCACTGACCATTCGCGAAGCCCTTGAAGAGGACACAGACGCTATTC 1285
|||||
QY 1021 TGCTTTGCTTCTTAACATCTATGGACAGATTTCGACTTCTGCTGAGATTTATATAGAAGGG 1080
|||||
Db 1286 TGCTTTGCTTCTTAACATCTATGGACAGATTTCGACTTCTGCTGAGATTTATATAGAAGGG 1345
|||||
QY 1081 GTTTCCTTCTTCTGACTCAGAGGCGACCCCTAAACAAGGAGAGATGAATCGAATCCAGAAG 1140
|||||
```

Db	1346	GT	TTCTCTCTCTGAC	TACAGAGCGGACCC	TACAAAGGAAGAGATGA	ATCGAATCCACAAG	1400	
Qy	1441	CC	AAATGAGGTGTCAT	CTCCTCCACTACCT	CTGCAAGTCATTCT	CTCAGCAGTACCCCAA	1200	
Db	1406	CC	AAATGAGGTGTCAT	CTCCTCCACTACCT	CTGCAAGTCATTCT	CTCAGCAGTACCCCAA	1465	
Qy	1201	GCC	CAGCATTTTGGTGG	CCCAACTCGTGTGG	CAACCATCCAGCAGT	GTCAAAAGCCCCACC	1360	
Db	1466	GCC	CAGCATTTTGGTGG	CCCAACTCGTGTGG	CAACCATCCAGCAGT	GTCAAAAGCCCCACC	1525	
Qy	1261	AA	TACTTTCGAGGAGT	TGGATGGAAACCTA	TCTTCGAGCTCTGT	TTTACAAAGATG	1320	
Db	1526	AA	TACTTTCGAGGAGT	TGGATGGAAACCTA	TCTTCGAGCTCTGT	TTTACAAAGATG	1585	
Qy	1321	CT	AAAAATTTGTGAGC	TTCTGAGGTGAGCT	TGGTGTCTTTTAATG	CAGAGTAAAAAGGA	1380	
Db	1586	CT	AAAAATTTGTGAGC	TTCTGAGGTGAGCT	TGGTGTCTTTGAATG	CAGAGTAAAAAGGA	1645	
Qy	1381	GCT	CCATCTCCTTAAG	TTGAGTGGTATAG	AAGAGGACTTTAATAG	AAGATTTCTCCAGAT	1440	
Db	1646	GCT	CCATCTCCTTAAG	TTGAGTGGTATAG	AAGAGGACTTTAATAG	AAGATTTCTCCAGAT	1705	
Qy	1441	TT	TAGGATTTTACAG	AAAAAACCTCGAT	CCATGGCAGAGCC	AGAGGAGATTTGCACCTTG	1500	
Db	1706	TT	TAGGATTTTACAG	AAAAAACCTCGAT	CCATGGCAGAGCC	AGAGGAGATTTGCACCTTG	1765	
Qy	1501	GT	CATTCTCGAGTGT	TTCAGAAATCTGG	TGCTTCACATGTACT	GCAGCAACAAA	1560	
Db	1766	GT	CATTCTCGAGTGT	TTCAGAAATCTGG	TGCTTCACATGTACT	GCAGCAACAAA	1825	
Qy	1561	TAC	GGCAGAGTGTCA	AGCATTGCGACGT	GCACTGCAGTGAG	AGAAATAGGACCTCAGCAAC	1620	
Db	1826	TAC	GGCAGAGTGTCA	AGCATTGCGACGT	GCACTGCAGTGAG	AGAAATAGGACCTCAGCAAC	1885	
Qy	1621	AAC	GGGTCTCTTCACT	CAGCCAACTCYACC	ACCACTTGGCAGCT	TATTTAGCCACAGCCC	1680	
Db	1886	AAC	GGGTCTCTTCACT	CAGCCAACTCTAC	CACTTGGCAGCT	TATTTAGCCACAGCCC	1945	
Qy	1681	TCCC	ACCCCACTCAGAG	CTCATCTGTGGA	CAACCCCAAACTCG	AGGGG	1740	
Db	1946	TCCC	ACCCCACTCAGAG	CTCATCTGTGGA	CAACCCCAAACTCG	AGGGG	2005	
Qy	1741	GTT	CTGTGTGAACCA	ATGAGCCCCGGT	CCAGCTCCAGATTTGG	CTTGTGTGCACTTC	1800	
Db	2006	GTT	CTGTGTGAACCA	ATGAGCCCCGGT	CCAGCTCCAGATTTGG	CTTGTGTGCACTTC	2065	
Qy	1801	AA	CTTGCTTGAAGAT	GACAAAGGAAGT	GAAGCATCTCCGAGG	CTGTGTGGTGACCAAC	1860	
Db	2066	GA	CTTGCTTGAAGAT	GACAAAGGAAGT	GAAGCATCTCCGAGG	CTGTGTGGTGACCAAC	2125	
Qy	1861	AG	CAGACAGGCGCAT	TCTTTSCAGGAG	GGTTCAACGGCAGG	CAACAAAAACCCCA	1920	
Db	2126	AG	CAGACAGGCGCAT	TCTTTSCAGGAG	GGTTCAACGGCAGG	CAACAAAAACCCCA	2185	
Qy	1921	GAG	CTTTTCCCGTGAA	AGAGCCCCCTCC	AGTTCTGGCCAA	ACCCAACTTGATTCC	1980	
Db	2186	GAG	CTTTTCCCGTGAA	AGAGCCCCCTCC	AGTTCTGGCCAA	ACCCAACTTGATTCC	2245	
Qy	1981	ACT	CAGTTTACAACAG	CTTCATACCAAGT	CTTACTTGGGAAC	AACCAACCAATTTGCAAAACCCCA	2040	
Db	2246	ACT	CAGTTTACAACAG	CTTCATACCAAGT	CTTACTTGGGAAC	AACCAACCAATTTGCAAAACCCCA	2305	
Qy	2041	CCT	CTTTCATCTCCT	TAAAGGATTTCT	CTTTCARCATGACT	GTTTTGA	CTCACTCAATGCTCCC	2100
Db	2306	CCT	CTTTCATCTCCT	TAAAGGATTTCT	CTTTCARCATGACT	GTTTTGA	CTCACTCAATGCTCCC	2365
Qy	2101	CC	AGCGGTGACAAC	ATCCARTAAGCAG	GTGAAGGCTCTT	CATCACAGACGTTTCAGCTTG	2160	
Db	2366	CC	AGCGGTGACAAC	ATCCARTAAGCAG	GTGAAGGCTCTT	CATCACAGACGTTTCAGCTTG	2425	
Qy	2161	GCC	GGCGGGAAGTAT	TTCTTCCCTCC	AGCAACACCA	CCGAGCACTGTGGCCCCCTTCC	2220	
Db	2426	GCC	GGCGGGAAGTAT	TTCTTCCCTCC	AGCAACACCA	CCGAGCACTGTGGCCCCCTTCC	2485	

[illegible]


```
QY 3301 CCCCCGAGCTGACATGGCTACTCAATGGCCAAACCTGTGCTACCAAGATGCTCCCAACAG 3360
Db 3566 CCCCCGAGCTGACATGGCTACTCAATGGCCAAACCTGTGCTACCAAGATGCTCCCAACAG 3625
QY 3361 ATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTACGCGGCAC 3420
Db 3626 ATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTACGCGGCAC 3685
QY 3421 GCAGGACCTATAAATGTCATCGCTACCAACAAACCGGCGAGAAATCTTTAGCTGGAG 3480
Db 3686 GCAGGACCTATAAATGTCATCGCTACCAACAAACCGGCGAGAAATCTTTAGCTGGAG 3745
QY 3481 CTCTCTGTAGTACCCAAAGAGTGTAAGAAAGCAACCTGTGATCCTCGGAGAACTACAGAAC 3540
Db 3746 CTCTCTGTAGTACCCAAAGAGTGTAAGAAAGCAACCTGTGATCCTCGGAGAACTACAGAAC 3805
QY 3541 TGGGTTGTTCCCGAAGCCACCCCGTGAGACTGGAGTGGCGGTGATAGCATGCCCCCA 3600
Db 3806 TGGGTTGTTCCCGAAGCCACCCCGTGAGACTGGAGTGGCGGTGATAGCATGCCCCCA 3865
QY 3601 CCTGTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGACACGAGAGAGATCAGT 3660
Db 3866 CCTGTGTTCTACTGGAAGAAAGCAATGAGACCATCCCTTGACACGAGAGAGATCAGT 3925
QY 3661 ATGACACGAGACACACAGGGTATGCTGCTCTCTCATTTGACCCAGCCCAAGAAATCAGAC 3720
Db 3926 ATGACACGAGACACACAGGGTATGCTGCTCTCTCATTTGACCCAGCCCAAGAAATCAGAC 3985
QY 3721 GCTGGATGTACAGTTGTCAGCCAGAAATGAAGCCGGCATCTGTCGTGCACTGCCAGG 3780
Db 3986 GCTGGATGTACAGTTGTCAGCCAGAAATGAAGCCGGCATCTGTCGTGCACTGCCAGG 4045
QY 3781 CTGGATATATACGCTCAGTGGCCACCATCAGATCCACGCCCATGCTGTCGCCGCCAGT 3840
Db 4046 CTGGATATATACGCTCAGTGGCCACCATCAGATCCACGCCCATGCTGTCGCCGCCAGT 4105
QY 3841 GCGAGTCGCTACGGATCTCTACCAAGTAAGGACTTTGACATATTTCTGCTTTTCTCTCC 3900
Db 4106 GCGAGTCGCTACGGATCTCTACCAAGTAAGGACTTTGACATATTTCTGCTTTTCTCTCC 4165
QY 3901 ATGGAAGCAGATGTGATTTATCATGCTCTTCGAGAGTAGTGAGAGTAGTGAACATT 3960
Db 4166 ATGGAAGCAGATGTGATTTATCATGCTCTTCGAGAGTAGTGAGAGTAGTGAACATT 4225
QY 3961 TAA 3963
Db 4226 TAA 4228

RESULT 4
HSM803687
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HSM803687 4683 bp mRNA linear PRI 10-JUL-2002
Homo sapiens mRNA; cDNA DKFZp451K027 (from clone DKFZp451K027).
AL832379
AL832379.1 GI:21732942
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Koehler,K., Beyer,A., Mewes,H.W., Well,B. and Wiemann,S.
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp451K027) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
```

```
FEATURES
Source
at http://mips.gsf.de/proj/cDNA/.
Location/Qualifiers
1..4683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp451K027"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal
polyA_site
4594
BASE COUNT 1376 a 1147 c 1048 g 1112 t
ORIGIN
Query Match 55.7%; Score 2602; DB 9; Length 4683;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3052; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 903 GTGGTACTGTGAAGCAAGGAGCTTGAATAATCCCCAGATATTCCACATCTCCAGGCAGG 962
Db 292 GTGGTACTGTGAAGCAAGGAGCTTGAATAATCCCCAGATATTCCACATCTCCAGGCAGG 351
QY 963 AAATCTGCACCTACTGACCAATTCGGGAAGCCTTTGAAGAGGACACAGGACGCTATTTCCTG 1022
Db 352 AAATCTGCACCTACTGACCAATTCGGGAAGCCTTTGAAGAGGACACAGGACGCTATTTCCTG 411
QY 1023 CTTTGGCTTCTAATCTATGGGACAGATTCGACTTTCGCTGAGATTTATATAGAAGGGT 1082
Db 412 CTTTGGCTTCTAATCTATGGGACAGATTCGACTTTCGCTGAGATTTATATAGAAGGGT 471
QY 1083 TTCTTCTTCTGACTCAGAAGCGCACCTAAAGGAAGAGATGAATCGAATCCAGAAGCC 1142
Db 472 TTCTTCTTCTGACTCAGAAGCGCACCTAAAGGAAGAGATGAATCGAATCCAGAAGCC 531
QY 1143 AAATGAGGTGTCTATCTCTCCCTACTACCTCTGCAGTCATTCTCCAGCAGTACCCCAAGC 1202
Db 532 AAATGAGGTGTCTATCTCTCCCTACTACCTCTGCAGTCATTCTCCAGCAGTACCCCAAGC 591
QY 1203 CCAGCATTTGGTGGCCCCAACCTCGTGTGGCAACCATCCAGCAGTGTCCAGAGCCCCACAA 1262
Db 592 CCAGCATTTGGTGGCCCCAACCTCGTGTGGCAACCATCCAGCAGTGTCCAGAGCCCCACAA 651
QY 1263 TTACTTTCAGGGGATTTGGATGGAACCTATCATTGACGCTCCTGTGTTTACAAGATGCT 1322
Db 652 TTACTTTCAGGGGATTTGGATGGAACCTATCATTGACGCTCCTGTGTTTACAAGATGCT 711
QY 1323 ACAAAATTTCTCAGCTTCTGAGGGTCAAGTGTGTTGCTTTTGAATGCAGAGTAAAGGAGC 1382
Db 712 ACAAAATTTCTCAGCTTCTGAGGGTCAAGTGTGTTGCTTTTGAATGCAGAGTAAAGGAGC 771
QY 1383 TCCATCTCTTAAGTTGAGTGGTATAGAGAGGAGCTTTAATAGAAGATTTCTCCAGATTT 1442
Db 772 TCCATCTCTTAAGTTGAGTGGTATAGAGAGGAGCTTTAATAGAAGATTTCTCCAGATTT 831
QY 1443 TAGGATTTTACAGAAAAACCTCGATCCATGGCAGAGCCAGAGAGATTTGCACCTTGGT 1502
Db 832 TAGGATTTTACAGAAAAACCTCGATCCATGGCAGAGCCAGAGAGATTTGCACCTTGGT 891
QY 1503 CATTGCTGAGGTGTTTTCAGAGAAGATTTCTGGTGTGCTTCCATGTACTGCAAGCAACAATA 1562
Db 892 CATTGCTGAGGTGTTTTCAGAGAAGATTTCTGGTGTGCTTCCATGTACTGCAAGCAACAATA 951
QY 1563 CGGCACAGTCTCAAGCATTCACAGCTGCACGTGAGAGGAAATAGAGACCTCAGCAACAA 1622
Db 952 CGGCACAGTCTCAAGCATTCACAGCTGCACGTGAGAGGAAATAGAGACCTCAGCAACAA 1011
QY 1623 CGGCTCTCTTCACTCAGCCAACTCYACCACCACTGCGAGCTATTGAGCCACAGCCCTC 1682
Db 1012 CGGCTCTCTTCACTCAGCCAACTCTACCAACCACTGGCAGCTATTGAGCCACAGCCCTC 1071
QY 1683 CCCACCCCACTCAGAGCTCCATCTGTGGAAACAAACCCCAACCCCAACCTCAGAGGGGT 1742
|||||
```

Db 1072 CCCACCCCACTAGAGCCTCCATCTGTGGAAACAACCCCAAAACCCCAAACTCGAGGGGT 1131
Qy 1743 TCTGGTGAACACAAATGAGCCCGGTCCAGCTCCAGGATGGGCTTCGTGTGCACCTTCAA 1802
Db 1132 TCTGGTGAACCAATGAGCCCGGTCCAGCTCCAGGATGGGCTTCGTGTGCACCTTCAA 1191
Qy 1803 CCTGCCCTGAAGATGACAAAGAAAGTGAAGCATCTCCGAGSGCTGGTGTGACCAACAG 1862
Db 1192 CCTGCCCTGAAGATGACAAAGAAAGTGAAGCATCTCCGAGSGCTGGTGTGACCAACAG 1251
Qy 1863 ACAGACAGGCCGATTTCTTTSCAGGAGAGTTCAACGGACAGGCAACAACACCCACAGA 1922
Db 1252 ACAGACAGGCCGATTTCTTTTCAAGGAGAGTTTCAACGGACAGGCAACAACACCCACAGA 1311
Qy 1923 GCCTTTCTTCCCGTGAAGAGCCCTCCAGTTCTGSCCAACCCCAAACTTGATTCAC 1982
Db 1312 GCCTTTCTTCCCGTGAAGAGCCCTCCAGTTCTGSCCAACCCCAAACTTGATTCAC 1371
Qy 1983 TCAGTTACAACAGCTTCATAACCAAGCTTTACTGGAACAACACCAATTTGCAAAACCCACC 2042
Db 1372 TCAGTTACAACAGCTTCATAAACAAGCTTTACTGGAACAACACCAATTTGCAAAACCCACC 1431
Qy 2043 TCCTTCATCTCTTAAGAGTTTCCTTTTCARCATGACTGTTTGAACCTCCAAATGCTCCCC 2102
Db 1432 TCCTTCATCTCTTAAGAGTTTCCTTTTCAGCATGACTGTTTGAACCTCCAAATGCTCCCC 1491
Qy 2103 AGCGGTGACAAACATCCARTAAGCAGGTGAAGGCTCCTTCATCACAGACGTTTCAGCTTGGC 2162
Db 1492 AGCGGTGACAAACATCCAGTAAGCAGGTGAAGGCTCCTTCATCACAGACGTTTCAGCTTGGC 1551
Qy 2163 CCGSCCGAAGTATTTCTTCCCTCCACGAACACCAACCGCAGCAACTGTGSCCCTTCCAG 2222
Db 1552 CCGSCCGAAGTATTTCTTCCCTCCACGAACACCAACCGCAGCAACTGTGSCCCTTCCAG 1611
Qy 2223 CTCTCCGGTGTCTACTTTGAGCAGCACTCCTCAAACTATTTCAGAGACAGTGAACAAGA 2282
Db 1612 CTCTCCGGTGTCTACTTTGAGCAGCACTCCTCAAACTATTTCAGAGACAGTGAACAAGA 1671
Qy 2283 AAGCCTCTTAGTGTCACACCCCTCTGTGCAACCAAACTCCTCAGAGGGCTTTCCATCCA 2342
Db 1672 AAGCCTCTTAGTGTCACACCCCTCTGTGCAACCAAACTCCTCAGAGGGCTTTCCATCCA 1731
Qy 2343 AAATGAGCCACTCCCAACGAGGCCCAACAGAACCCACCAACCACTTCACATTTTCCAT 2402
Db 1732 AAATGAGCCACTCCCAACGAGGCCCAACAGAACCCACCAACCACTTCACATTTTCCAT 1791
Qy 2403 CCCCAGCGGAACACGATTTGAGCCCCCTGTGTGTCGCCCAATTCCTGTCTCTCTACACAG 2462
Db 1792 CCCCAGCGGAACACGATTTGAGCCCCCTGTGTGTCGCCCAATTCCTGTCTCTCTACACAG 1851
Qy 2463 CCGGATTCAGAACCCAGTGGCTTTCCCTCAGCTCTGTGCTGTGCTCTCCTCGCATCCC 2522
Db 1852 CCGGATTCAGAACCCAGTGGCTTTCCCTCAGCTCTGTGCTGTGCTCTCCTCGCATCCC 1911
Qy 2523 ACCACAAAATGCCATGGRGCTGTAGAAGTGCACCACTCCATGCCATCCCAGGAGTTAGC 2582
Db 1912 ACCACAAAATGCCATGGRGCTGTAGAAGTGCACCACTCCATGCCATCCCAGGAGTTAGC 1971
Qy 2583 GAAGAAAAATACAAGTCTCCTCAACCAGTGAATGATGATAACATTCGTTGAACACTAGAA 2642
Db 1972 GAAGAAAAATACAAGTCTCCTCAACCAGTGAATGATGATAACATTCGTTGAACACTAGAA 2031
Qy 2643 CGCAGTATTCGAGACTTGGGAAAAAATAAATTTTCAGTGTGATGATGATGATGATGATGAT 2702
Db 2032 CGCAGTATTCGAGACTTGGGAAAAAATAAATTTTCAGTGTGATGATGATGATGATGATGAT 2091
Qy 2703 GGAGTACAAAATTTCAAGCTTTGAGCAGGCGTGAATGAATGAATGAATGAATGAATGAAT 2762
Db 2092 GGAGTACAAAATTTCAAGCTTTGAGCAGGCGTGAATGAATGAATGAATGAATGAATGAAT 2151
Qy 2763 ACGTACTCCTTTGATGAATCAGATGATGAATTCACATGATGATGATGATGATGATGATGAT 2822
Db 2152 ACGTACTCCTTTGATGAATCAGATGATGAATTTCAACATGATGATGATGATGATGATGATGAT 2211

Qy 2823 GTGTATTGCTCCCATCTTTTGACAAGAGACTCAAGCACTTCGGGCTCACAGAAGCTCTCC 2882
Db 2212 GTGTATTGCTCCCATCTTTTGACAAGAGACTCAAGCACTTCGGGCTCACAGAAGCTCTCC 2271
Qy 2883 AGTCACATTCACCTGCAAAATTTGTTGGGATACCTGTTTCCAAAGGTTTACTGGTTCAAAGA 2942
Db 2272 AGTCACATTCACCTGCAAAATTTGTTGGGATACCTGTTTCCAAAGGTTTACTGGTTCAAAGA 2331
Qy 2943 TGGGAACAGATTTCTTAAGAGAAATGAGCACTGCAAAATGAGGCGAGAAGAGATGGAC 3002
Db 2332 TGGGAACAGATTTCTTAAGAGAAATGAGCACTGCAAAATGAGGCGAGAAGAGATGGAC 2391
Qy 3003 ATGCTCTCTGCACATTCGAATCCACTACAGTGTAGCAGTGGCAACTACACCATCATGGC 3062
Db 2392 ATGCTCTCTGCACATTCGAATCCACTACAGTGTAGCAGTGGCAACTACACCATCATGGC 2451
Qy 3063 AGCCAACCCCAAGGGAGAAATCAGCTGTTCTGGCCACTTGTATGGTACAAAATTTGCCCAT 3122
Db 2452 AGCCAACCCCAAGGGAGAAATCAGCTGTTCTGGCCACTTGTATGGTACAAAATTTGCCCAT 2511
Qy 3123 TCGCAGTTCGGTAACTCTGCTGCTCAGTCTCAGAGGGGAAGATCCCAGATGCAAGAAAG 3182
Db 2512 TCGCAGTTCGGTAACTCTGCTGCTCAGTCTCAGAGGGGAAGATCCCAGATGCAAGAAAG 2571
Qy 3183 AGCAAAAGAGCCCTCAGGAACGCTTTTTCGAGCACATTTTCTCTGAGGCTCCTGGGA 3242
Db 2572 AGCAAAAGAGCCCTCAGGAACGCTTTTTCGAGCACATTTTCTCTGAGACTCCTGGGA 2631
Qy 3243 TATGGTAGCTCATGAGGGGCGCTCTCTCGGCTGGACTTAAGGTGAGTGTGTACCGCC 3302
Db 2632 TATGGTAGCTCATGAGGGGCGCTCTCTCGGCTGGACTTAAGGTGAGTGTGTACCGCC 2691
Qy 3303 CCGGAGCTGACATGGCTACTCAATGGCCAACTGTGCTACCAGATGCTCCCAACAGAT 3362
Db 2692 CCGGAGCTGACATGGCTACTCAATGGCCAACTGTGCTACCAGATGCTCCCAACAGAT 2751
Qy 3363 GCTGCTCAGGAGACCGGAGTCCACTCTCTCATTTGACCCACTCACTCAGCGCAGCG 3422
Db 2752 GCTGCTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCGCAGCG 2811
Qy 3423 AGGACCTATAAGTGCATCGCTACCAAAACCGGCGAGAAATTTCTTTAGCTGGAGCT 3482
Db 2812 AGGACCTATAAGTGCATCGCTACCAAAACCGGCGAGAAATTTCTTTAGCTGGAGCT 2871
Qy 3483 CTCTGTAGTAGCCAAAGAGGTGAAGAACCACTGTGATCCTGAGAAACTACAGAACTG 3542
Db 2872 CTCTGTAGTAGCCAAAGAGGTGAAGAACCACTGTGATCCTGAGAAACTACAGAACTG 2931
Qy 3543 CGGTGTTCCCGAAGGCCACCCGCTGAGACTGGAGTGGCGGTGATAGGCATGCCCCACC 3602
Db 2932 CGGTGTTCCCGAAGGCCACCCGCTGAGACTGGAGTGGCGGTGATAGGCATGCCCCACC 2991
Qy 3603 TGTGTTCTACTGGAAGAAAGACAAATGAGACATCCCTTTGCACCAGAGAGGATCAGTAT 3662
Db 2992 TGTGTTCTACTGGAAGAAAGACAAATGAGACATCCCTTTGCACCAGAGAGGATCAGTAT 3051
Qy 3663 GCACCAGGACACACAGGGTATGCCTGCCCTTCTCATTTACGCCAGCCAAAGAAATCAGACGC 3722
Db 3052 GCACCAGGACACACAGGGTATGCCTGCCCTTCTCATTTACGCCAGCCAAAGAAATCAGACGC 3111
Qy 3723 TGGATGTACACGTTGTTCAGCAAGAAATGAAGCGGATCGTGTGCTGCACTGCCAGGCT 3782
Db 3112 TGGATGTACACGTTGTTCAGCAAGAAATGAAGCGGATCGTGTGCTGCACTGCCAGGCT 3171
Qy 3783 GGATATATACGCTAGTGGCAACCATCAGATCCCAACGCCCATGTCTGTCCGGGCCAGTGG 3842
Db 3172 GGATATATACGCTAGTGGCAACCATCAGATCCCAACGCCCATGTCTGTCCGGGCCAGTGG 3231
Qy 3843 CAGTCCGTACGGATCTCTCACCAGTAAAGNACTTGACATATTTCTGCCCTTTTCTCCTCAT 3902
Db 3232 CAGTCCGTACGGATCTCTCACCAGTAAAGNACTTGACATATTTTCTGCCCTTTTCTCCTCAT 3291

QY 3903 GGAACACGATGGTGTATTGCTCTTCTCGAGTGTAGTGGAGGATGATGAACCTTTA 3962
|||||
Db 3292 GGAACACGATGGTGTATTGCTCTTCTCGAGTGTAGTGGAGGATGATGAACCTTTA 3351
QY 3963 A 3963
Db 3352 A 3352

RESULT 5
AK027343
LOCUS
DEFINITION
Homo sapiens cDNA FLJ14437 fis, clone HEMB1000083, weakly similar
to MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES
(EC 2.7.1.117).

ACCESSION
AK027343
VERSION
AK027343.1 GI:14041954
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens embryo, 10 weeks whole embryo, mainly body cDNA to
mRNA, clone_lib:HEMB1 clone:HEMB1000083.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2127)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction;
5'- & 3'-end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source
1..2127
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMB1000083"
/tissue_type="whole embryo, mainly body"
/clone_lib="HEMB1"
/dev_stage="embryo, 10 weeks"
/note="cloning vector: pME185FL3"
231..1754
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55048.1"
/db_xref="GI:14041955"
translation="*WDDSIKASTISQILLRESYLAETRRHGNERSRAEPPSNPCHF
GSPGAEGGGQDLDLPSAFLSQELDESYNLARINAYDPLKADETQAKRLSP
DOMKLSPNLSFPNFCQDNPRSPTSKSPQAKRPOYCSQTSKRVFLNKADRIIE
LSSLFKSHSKRIRPRACKNHSKLESQNKVQENSFSFSDLSERRSSVPIPAD
TRDENVNHALEQOAKRREAQEAASEAGGDTPGSSSSLIYEPLGPPPTOKLR
SREVEPTGRVQDCLVIGLPPQVRWYCEGKELENSPDIIHVOAGNLHSLTIAFAEE
DFGRVSCFASNIYGTDSAEYIEGVSSSDSEGPNKEEMNRIOKPNRSPPTTSA
VTPPAVQAHLVAPRVATIQCCSPNLYLQLDGKPIIAAPVFTKMLQNLASGEQ
LVVFECRVKGAPSPRVWYREGLTIEDSPDFILQKKPRMAEPCKDDFNFLLYE"

BASE COUNT
638 a 501 c 487 g 501 t
ORIGIN

Query Match
37.4%; Score 1483; DB 9; Length 2127;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCAAGACACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAAGAGAGCTAT 60
|||||
Db 231 ATCAAGACACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAAGAGAGCTAT 290
QY 61 TTAGCTGAAACACAGACATCGGGGAAACAATGAGAGGAGTGCAGCGGAGCCCTCCTCCAAC 120
|||||
Db 291 TTAGCTGAAACACAGACATCGGGGAAACAATGAGAGGAGTGCAGCGGAGCCCTCCTCCAAC 350
QY 121 CTTTGGCCATTTCCGGCAGTCCCTTCTGGGGCCGCTGAAGAGGGCGGAGGCCAAGATGACCTT 180
|||||
Db 351 CTTTGGCATTTCGGCAGTCCCTTCTGGGGCCGCTGAAGAGGGCGGAGGCCAAGATGACCTT 410
QY 181 CCAGATCTTTTCAGCCTTTCTGAGCCAAAGAAGAAATAGACCAAAAGTGTCATTTGGCAAGA 240
|||||
Db 411 CCAGATCTTTTCAGCCTTTCTGAGCCAAAGAAGAAATAGACCAAAAGTGTCATTTGGCAAGA 470
QY 241 CTGGCCATCAATTTACGACCCCTTTGGAGAAAGGCAGATGAAACTCAAGCTAGAAAACGACTT 300
Db 471 CTGGCCATCAATTTACGACCCCTTTGGAGAAAGGCAGATGAAACTCAAGCTAGAAAACGACTT 530
QY 301 TCTCTGATCAGATGAAACACTCACCTAATTTAAGTTTTCAGCCTAACTTCTGCCAGGAT 360
Db 531 TCTCTGATCAGATGAAACACTCACCTAATTTAAGTTTTCAGCCTAACTTCTGCCAGGAT 590
QY 361 AACCTTCGAAGTCCCAAGCAGCTCTAAGAAGCCGCCAGGAGGCAAAAAGGCCACAGTAT 420
Db 591 AACCTTCGAAGTCCCAAGCAGCTCTAAGAAGCCGCCAGGAGGCAAAAAGGCCACAGTAT 650
QY 421 TGTCTTGAACACCCAGTCCCAAAAAGTATTTTAAATAAGGCTGCCGACTTCATTGAAGAG 480
Db 651 TGTCTTGAACACCCAGTCCCAAAAAGTATTTTAAATAAGGCTGCCGACTTCATTGAAGAG 710
QY 481 CTATCTCTCCCTTTTCAAAATCCCAAGCTCCCAAGGATTAGACTCTGCTGCTGCAAAAAC 540
Db 711 CTATCTCTCCCTTTTCAAAATCCCAAGCTCCCAAGGATTAGACTCTGCTGCTGCAAAAAC 770
QY 541 CACAAGAGTAAACTGGAATCTCAAAAAGTATTCAGCAAAAACAGCTCCAGTTTCTCA 600
Db 771 CACAAGAGTAAACTGGAATCTCAAAAAGTATTCAGCAAAAACAGCTCCAGTTTCTCA 830
QY 601 GATCTGTCAAAAAGACAGAAAGATCTTCTGTTCCCATCCCTATPCCCTGGGATACCCAGG 660
Db 831 GATCTGTCAAAAAGACAGAAAGATCTTCTGTTCCCATCCCTATPCCCTGGGATACCCAGG 890
QY 661 GATAATGAAGTGAATCACGCCCTGGAAACACAGCAAGCAAGAGGCGTGAAGCGGAGCAG 720
Db 891 GATAATGAAGTGAATCACGCCCTGGAAACACAGCAAGCAAGAGGCGTGAAGCGGAGCAG 950
QY 721 GCTGCCAGTCCAGGCGCTGGTGGAGACACTACACAGGGTCTTCCCTTCATCTCTGTAC 780
Db 951 GCTGCCAGTCCAGGCGCTGGTGGAGACACTACACAGGGTCTTCCCTTCATCTCTGTAC 1010
QY 781 TATGAAGAACTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGGAGCAGAGAAGTT 840
Db 1011 TATGAAGAACTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGGAGCAGAGAAGTT 1070
QY 841 CCAGAAGAACTCCAGTACAGTTGGATTGATAGTGGTAGGAATTCACCACCTCAAGTA 900
Db 1071 CCAGAAGAACTCCAGTACAGTTGGATTGATAGTGGTAGGAATTCACCACCTCAAGTA 1130
QY 901 AGCTGTACTGTCAAGCAAGGAGCTTGAANAATCCCAATATTCACATGCTCCAGGCA 960
Db 1131 AGCTGTACTGTCAAGCAAGGAGCTTGAANAATCCCAATATTCACATGCTCCAGGCA 1190
QY 961 GAAATCTGCACCTCACTGACCATTTGCCGAAGCCCTTTGAAGAGGACACAGACGCTATTC 1020
Db 1191 GAAATCTGCACCTCACTGACCATTTGCCGAAGCCCTTTGAAGAGGACACAGACGCTATTC 1250
QY 1021 TGCTTTGCTTCTAACATCTATGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1080
|||||

Db 1251 TGCTTTGCTTCTACATCATATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG 1310
Qy 1081 GTTCTCTTCTGACTCAGAAAGGACCTAACAAAGGAGAGATGAATCGAATCCAGAAG 1140
Db 1311 GTTCTCTTCTGACTCAGAAGGCGACCTAACAAAGGAGAGATGAATCGAATCCAGAG 1370
Qy 1141 CCAATAGAGGTGCATCTCCCTCCACTACCTCTCTGAGTTCCTCCAGCAGTACCCCAA 1200
Db 1371 CCAATAGAGGTGCATCTCCCTCCACTACCTCTCTGAGTTCCTCCAGCAGTACCCCAA 1430
Qy 1201 GCCAGCATTTGGTGGCCCACTCGTGTGCAACCATCCAGAGTGTGAGAGCCCCACC 1260
Db 1431 GCCAGCATTTGGTGGCCCACTCGTGTGCAACCATCCAGAGTGTGAGAGCCCCACC 1490
Qy 1261 AATTACTTGCAGGATTTGGATGGAACCTATCATTTGCAGCTCTCTGTTTACAAAAGATG 1320
Db 1491 AATTACTTGCAGGATTTGGATGGAACCTATCATTTGCAGCTCTCTGTTTACAAAAGATG 1550
Qy 1321 CTACAAAATTTGTGAGCTTCTGAGGGTCAGCTGGTGTGTTTGAATGCAGAGTAAAAGGA 1380
Db 1551 CTACAAAATTTGTGAGCTTCTGAGGGTCAGCTGGTGTGTTTGAATGCAGAGTAAAAGGA 1610
Qy 1381 GCTCCATCTCTAAGTGTGAGTGTATAGAAAGGGACTTTAATAGAGATTTCTCCAGAT 1440
Db 1611 GCTCCATCTCTAAGTGTGAGTGTATAGAAAGGGACTTTAATAGAGATTTCTCCAGAT 1670
Qy 1441 TTTAGGATTTTACAGAAAAACCTCGATCCATGGCAGAGCCAG 1483
Db 1671 TTTAGGATTTTACAGAAAAACCTCGATCCATGGCAGAGCCAG 1713

RESULT 6
AC024258 160658 bp DNA linear PRI 14-MAY-2002
LOCUS Homo sapiens chromosome 10 clone RP11-297N15, complete sequence.
AC024258
VERSION AC024258.9 GI:20564414
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 160658)
Direct Submission
Submitted (28-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
3 (bases 1 to 160658)
Smith,D.R.
Direct Submission
Submitted (13-NOV-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
4 (bases 1 to 160658)
Smith,D.R.
Direct Submission
Submitted (14-MAY-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 14, 2002 this sequence version replaced gi:16905139.
FEATURES
source
1..160658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-297N15"
/clone_1lb="pCI-11"

BASE COUNT 50744 a 31675 c 30448 g 47791 t
ORIGIN

Query Match 22.8%; Score 904; DB 9; Length 160658;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAAGACGACAGATAGAAAGCTTCTACTTCCATATCTCAAGCTTCTTAAGAGAGAGCTAT 60
Db 110106 ATGCAAGACGACAGATAGAAAGCTTCTACTTCCATATCTCAAGCTTCTTAAGAGAGAGCTAT 110165
Qy 61 TTAGCTGAAACCCACACATCGGGGAAACAATAGAGAGAGTTCGAGCGGAGCCCTCTCCAAAC 120
Db 110166 TTAGCTGAAACCCACACATCGGGGAAACAATAGAGAGAGTTCGAGCGGAGCCCTCTCCAAAC 110225
Qy 121 CTTTGCCATTTTCGGCAGTCCCTTCTGGGGCCGCTGAAGGAGGCGGAGGCAAGATGACCTT 180
Db 110226 CTTTGCCATTTTCGGCAGTCCCTTCTGGGGCCGCTGAAGGAGGCGGAGGCAAGATGACCTT 110285
Qy 181 CCAGATCTTTCAGCCTTTCTGAGCCAAAGAAATTTAGACGAAAGTGCAATTTTGGCAAGA 240
Db 110286 CCAGATCTTTCAGCCTTTCTGAGCCAAAGAAATTTAGACGAAAGTGCAATTTTGGCAAGA 110345
Qy 241 CTGGCCATCAATTAAGACCTTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAAACGACTT 300
Db 110346 CTGGCCATCAATTAAGACCTTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAAACGACTT 110405
Qy 301 TCTCCTGATCAGATGAACACTCACCTTAATTTAAAGTTTTTGAGCCTAACTTCTGCCAGAT 360
Db 110406 TCTCCTGATCAGATGAACACTCACCTTAATTTAAAGTTTTTGAGCCTAACTTCTGCCAGAT 110465
Qy 361 AACCTCGAAGTCCCAAGCTCTTAAGAAAGCCCCCAGAGGCAAAAAGGCCACAGTAT 420
Db 110466 AACCTCGAAGTCCCAAGCTCTTAAGAAAGCCCCCAGAGGCAAAAAGGCCACAGTAT 110525
Qy 421 TGTTCGAAACCCAGTCCAAAAAGATTTTAAATAGGTCGCGACTTCATTGAAGAG 480
Db 110526 TGTTCGAAACCCAGTCCAAAAAGATTTTAAATAGGTCGCGACTTCATTGAAGAG 110585
Qy 481 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAAGGATTAGACCTCGTGCCTGCAAAAAC 540
Db 110586 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAAGGATTAGACCTCGTGCCTGCAAAAAC 110645
Qy 541 CACAAGAGTAACTGGGAATCTCAAAACAAGTTTATGAGGAAAAACAGCTCCAGTTTCTCA 600
Db 110646 CACAAGAGTAACTGGGAATCTCAAAACAAGTTTATGAGGAAAAACAGCTCCAGTTTCTCA 110705
Qy 601 GATCTGTCAAGAACACGAGAAAGATCTTCTTCCCATCCCTATCCCTGCGGATACACAGG 660
Db 110706 GATCTGTCAAGAACACGAGAAAGATCTTCTTCCCATCCCTATCCCTGCGGATACACAGG 110765
Qy 661 GATAATGAAGTGAATCAGCCCTCGAAGACAGAGGAAAGGCGGTGAAGCGGAGCAG 720
Db 110766 GATAATGAAGTGAATCAGCCCTCGAAGACAGAGGAAAGGCGGTGAAGCGGAGCAG 110825
Qy 721 GCTGCCAGTGAAGCGGCTGGTGGAGACACTACACAGGAGTCTCCCTTCATCTCTGTAC 780
Db 110826 GCTGCCAGTGAAGCGGCTGGTGGAGACACTACACAGGAGTCTCCCTTCATCTCTGTAC 110885
Qy 781 TATGAAGAACCTCTGGGGCAACCTCCCGGTTCACTCAAAAAGTTTACGAGCAGAGAGAGTT 840
Db 110886 TATGAAGAACCTCTGGGGCAACCTCCCGGTTCACTCAAAAAGTTTACGAGCAGAGAGAGTT 110945
Qy 841 CCAGAAGGAACCTCGAGTACAGTTGGATTGTCATAGTGTAGGAAATCCACCACCTCAAGTA 900
Db 110946 CCAGAAGGAACCTCGAGTACAGTTGGATTGTCATAGTGTAGGAAATCCACCACCTCAAGTA 111005
Qy 901 AGGT 904
Db 111006 AGGT 111009

RESULT 7
AC063964 175265 bp DNA linear HTG 30-AUG-2001
LOCUS Homo sapiens chromosome 10 clone RP11-31303, WORKING DRAFT
DEFINITION

```

SEQUENCE, 20 unordered pieces.
AC063964
VERSION AC063964.3 GI:9897616
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 175265)
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
REFERENCE 2 (bases 1 to 175265)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (22-APR-2000) Genome Therapeutics Corporation, 100 Beaver
AUTHORS Street, Waltham, MA 02453, USA
JOURNAL On Aug 24, 2000 this sequence version replaced gi:8439815.
COMMENT ----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg198
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 153700 bases at least Q40
Consensus quality: 164398 bases at least Q30
Consensus quality: 167130 bases at least Q20
Insert size: 173464; sum-of-contigs
Quality coverage: 4.1x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1145: contig of 1145 bp in length
* 1146 1245: gap of unknown length
* 1246 2414: contig of 1169 bp in length
* 2415 2514: gap of unknown length
* 2515 3864: contig of 1350 bp in length
* 3865 3964: gap of unknown length
* 3965 5078: contig of 1114 bp in length
* 5079 5178: gap of unknown length
* 5179 6647: contig of 1469 bp in length
* 6648 7927: gap of unknown length
* 7928 8026: contig of 1179 bp in length
* 8027 9412: contig of 1386 bp in length
* 9413 9512: gap of unknown length
* 9513 10563: contig of 1051 bp in length
* 10564 10663: gap of unknown length
* 10664 11995: contig of 1332 bp in length
* 11996 12095: gap of unknown length
* 12096 13309: contig of 1214 bp in length
* 13310 13409: gap of unknown length
* 13410 15172: contig of 1763 bp in length
* 15173 15272: gap of unknown length
* 15273 17973: contig of 2701 bp in length
* 17974 18073: gap of unknown length
* 18074 21419: contig of 3346 bp in length
* 21420 24881: gap of unknown length
* 24882 24981: gap of unknown length
* 24982 31876: contig of 6895 bp in length

```

```

* 31877 31976: gap of unknown length
* 31977 49015: contig of 17039 bp in length
* 49016 49115: gap of unknown length
* 49116 66286: contig of 17171 bp in length
* 66287 66387: gap of unknown length
* 66388 86201: contig of 19815 bp in length
* 86202 86302: gap of unknown length
* 86303 132480: contig of 46179 bp in length
* 132481 132581: gap of unknown length
* 132581 175265: contig of 42685 bp in length.
FEATURES
Location/Qualifiers
Source
1..175265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-31303"
/clone_lib="RPC1-11"
1..1145
/note="assembly_name:Contig1"
misc_feature 1246..2414
/note="assembly_name:Contig11"
misc_feature 2515..3864
/note="assembly_name:Contig12"
misc_feature 3965..5078
/note="assembly_name:Contig15"
misc_feature 5179..6647
/note="assembly_name:Contig18"
misc_feature 6748..7926
/note="assembly_name:Contig20"
misc_feature 8027..9412
/note="assembly_name:Contig23"
misc_feature 9513..10563
/note="assembly_name:Contig25"
misc_feature 10664..11995
/note="assembly_name:Contig27"
misc_feature 12096..13309
/note="assembly_name:Contig28"
misc_feature 13410..15172
/note="assembly_name:Contig29"
misc_feature 15273..17973
/note="assembly_name:Contig30"
misc_feature 18074..21419
/note="assembly_name:Contig31"
misc_feature 21520..24881
/note="assembly_name:Contig32"
misc_feature 24982..31876
/note="assembly_name:Contig33"
misc_feature 31977..49015
/note="assembly_name:Contig34"
misc_feature 49116..66286
/note="assembly_name:Contig35"
clone_end:SP6"
misc_feature 66387..86201
/note="assembly_name:Contig36"
misc_feature 86302..132480
/note="assembly_name:Contig37"
clone_end:T7"
misc_feature 132581..175265
/note="assembly_name:Contig38"
BASE COUNT 53132 a 34451 c 33650 g 52102 t 1930 others
ORIGIN
Query Match 22.8%; Score 904; DB 2; Length 175265;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCAACAGCAGACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 60
|||||
Db 162072 ATGCAACAGCAGACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 162131
QY 61 TTAGCTGAAACCCAGACATCGGGGAAACAATGAGAGGAGTGCAGCGGAGCCCTCTCCCAAC 120
|||||
Db 162132 TTAGCTGAAACCCAGACATCGGGGAAACAATGAGAGGAGTGCAGCGGAGCCCTCTCCCAAC 162191

```

```
Oy 121 CTTTGCCATTTTCGCGAGTCTTCTGGGCCGCTGAAGGAGCGGAGGCGCAAGATGACCTT 180
|||||
Db 162192 CTTTGCCATTTTCGCGAGTCTTCTGGGCCGCTGAAGGAGCGGAGGCGCAAGATGACCTT 162251
|||||
Oy 181 CCAGATCTTTCAGGCTTCTGAGCAGAGAGAAATAGACGAAAGTGCAATTTGGCAGA 240
|||||
Db 162252 CCAGATCTTTCAGGCTTCTGAGCAGAGAGAAATAGACGAAAGTGCAATTTGGCAGA 162311
|||||
Oy 241 CTGGCCATCAATTAGGACCTTTGAGAGAGCGACATGAACCTCAAGCTAGAAAAAGGACTT 300
|||||
Db 162312 CTGGCCATCAATTAGGACCTTTGAGAGAGCGACATGAACCTCAAGCTAGAAAAAGGACTT 162371
|||||
Oy 301 TCTCCTGATCAGATGAACACTCACCTCAATTTAAGTTTTCGAGCCTTAAGTTCTGCCAGGAT 360
|||||
Db 162372 TCTCCTGATCAGATGAACACTCACCTCAATTTAAGTTTTCGAGCCTTAAGTTCTGCCAGGAT 162431
|||||
Oy 361 AACCTTGAAGTCCGACAGCTCTAAGAAAGCCCCAGGAGGCAAAAGGCCACAGTAT 420
|||||
Db 162432 AACCTTGAAGTCCGACAGCTCTAAGAAAGCCCCAGGAGGCAAAAGGCCACAGTAT 162491
|||||
Oy 421 TGTTCCTGAACCCAGTCCAAAAAGTATTTTAAATAGGCTGCCGACTTCATTGAAGAG 480
|||||
Db 162492 TGTTCCTGAACCCAGTCCAAAAAGTATTTTAAATAGGCTGCCGACTTCATTGAAGAG 162551
|||||
Oy 481 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAAGGATTAGACCTCGTGCCTGCAAAAAC 540
|||||
Db 162552 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAAGGATTAGACCTCGTGCCTGCAAAAAC 162611
|||||
Oy 541 CACAAGAGTAACTGGAACTCTCAAAACAAGATTATGAGGAAACAGCTCCAGTTCTCTCA 600
|||||
Db 162612 CACAAGAGTAACTGGAACTCTCAAAACAAGATTATGAGGAAACAGCTCCAGTTCTCTCA 162671
|||||
Oy 601 GATCTGTGAGAAACAGAGAAAGATCTCTGTCCCATCCCTATCCCTGCGGATACACAG 660
|||||
Db 162672 GATCTGTGAGAAACAGAGAAAGATCTCTGTCCCATCCCTATCCCTGCGGATACACAG 162731
|||||
Oy 661 GATAATGAAGTGAATCAGCCCTCGAAGCAGCAGAGGAGGAGGCGTGAAGCGGAGCAG 720
|||||
Db 162732 GATAATGAAGTGAATCAGCCCTGGAACAGCAGAGGAGGAGGCGTGAAGCGGAGCAG 162791
|||||
Oy 721 GCTGCCAGTGAAGCGGCTGTGGAGACACTACACAGGAGTCTCCCTTCATCTCTCTAC 780
|||||
Db 162792 GCTGCCAGTGAAGCGGCTGTGGAGACACTACACAGGAGTCTCCCTTCATCTCTCTAC 162851
|||||
Oy 781 TATGAAGAACTCTGGGGCAACCTCCCGGTTCACTCAAAAGTTACGAGCAGAGAAAGTT 840
|||||
Db 162852 TATGAAGAACTCTGGGGCAACCTCCCGGTTCACTCAAAAGTTACGAGCAGAGAAAGTT 162911
|||||
Oy 841 CCAGAAGAACTCGAGTACAGTTGGATTGTCATAGTGTAGGAATTCACCACCTCAAGTA 900
|||||
Db 162912 CCAGAAGAACTCGAGTACAGTTGGATTGTCATAGTGTAGGAATTCACCACCTCAAGTA 162971
|||||
Oy 901 AGGT 904
|||||
Db 162972 AGGT 162975

RESULT 8
AC024460/c 180707 bp DNA linear HTG 28-MAR-2000
LOCUS Homo sapiens clone RP11-287P14, WORKING DRAFT SEQUENCE, 22
DEFINITION unordered pieces.
ACCESSION AC024460
VERSION AC024460.2 GI:7331497
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180707)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-287P14
```

JOURNAL
REFERENCE
AUTHORSUnpublished
2 (bases 1 to 180707)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavskiy,L.,
Boukigaiter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grant-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneau,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,
Petersen,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 28, 2000 this sequence version replaced gi:7108256.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7059

Center clone name: 287_P_14

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 168456 bases at least Q40

Consensus quality: 174769 bases at least Q30

Consensus quality: 176914 bases at least Q20

Insert size: 188000; agarose-fp

Insert size: 178607; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 1528: contig of 1528 bp in length
* 1529 1628: gap of 100 bp
* 1629 2914: contig of 1286 bp in length
* 2915 3014: gap of 100 bp
* 3015 5061: contig of 2047 bp in length
* 5062 5161: gap of 100 bp
* 5162 7974: contig of 2813 bp in length
* 7975 8074: gap of 100 bp
* 8075 11732: contig of 3658 bp in length
* 11733 11832: gap of 100 bp
* 11833 14787: contig of 2955 bp in length
* 14788 14887: gap of 100 bp
* 14888 19746: contig of 4859 bp in length
* 19747 19846: gap of 100 bp
* 19847 23870: contig of 4024 bp in length

```
* 23871 23970: gap of 100 bp
* 23971 28748: contig of 4778 bp in length
* 28749 28848: gap of 100 bp
* 28849 35292: contig of 6444 bp in length
* 35293 35392: gap of 100 bp
* 35393 40557: contig of 4965 bp in length
* 40558 40457: gap of 100 bp
* 40458 45936: contig of 5479 bp in length
* 45937 46036: gap of 100 bp
* 46037 52180: contig of 6144 bp in length
* 52181 52280: gap of 100 bp
* 52281 57511: contig of 5231 bp in length
* 57512 57611: gap of 100 bp
* 57612 65019: contig of 7408 bp in length
* 65020 65119: gap of 100 bp
* 65120 73479: contig of 8360 bp in length
* 73480 73579: gap of 100 bp
* 73580 81893: contig of 8314 bp in length
* 81894 81993: gap of 100 bp
* 81994 92063: contig of 10070 bp in length
* 92064 92163: gap of 100 bp
* 92164 103396: contig of 11233 bp in length
* 103397 103496: gap of 100 bp
* 103497 112935: contig of 9439 bp in length
* 112936 113035: gap of 100 bp
* 113036 126895: contig of 13860 bp in length
* 126896 126995: gap of 100 bp
* 126996 180707: contig of 53712 bp in length.

FEATURES
    source
    1. .180707
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="RP11-287P14"
        /clone_lib="RP11-11 Human Male BAC"
    1. .1528
        /note="assembly_fragment"
        1629. .2914
        /note="assembly_fragment"
        3015. .5061
        /note="assembly_fragment"
        5162. .7974
        /note="assembly_fragment"
        8075. .11732
        /note="assembly_fragment"
        11833. .14787
        /note="assembly_fragment"
        14888. .19746
        /note="assembly_fragment"
        19847. .23870
        /note="assembly_fragment"
        23971. .28748
        /note="assembly_fragment"
        28849. .35292
        /note="assembly_fragment"
        35393. .40357
        /note="assembly_fragment"
        40458. .45936
        /note="assembly_fragment"
        46037. .52180
        /note="assembly_fragment"
        52281. .57511
        /note="assembly_fragment"
        57612. .65019
        /note="assembly_fragment"
        65120. .73479
        /note="assembly_fragment"
        73580. .81893
        /note="assembly_fragment"
        81994. .92063
        /note="assembly_fragment"
        92164. .103396
        /note="assembly_fragment"
        103497. .112935
```

```
/note="assembly_fragment"
clone_end:T7
vector_side:left"
113036. .126895
/note="assembly_fragment"
126996. .180707
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
BASE COUNT 55951 a 33545 c 34680 g 54429 t 2102 others
ORIGIN
Query Match 22.8%; Score 904; DB 2; Length 180707;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAAGACGACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAAGAGAGAGCTAT 60
|||||
Db 154380 ATGCAAGACGACAGCATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAAGAGAGAGCTAT 154321
|||||
QY 61 TTAGCTGAACACGACATCGGGAAACAATGAGAGAGTTCGAGGGAGCCCTCTCTCAAC 120
|||||
Db 154320 TTAGCTGAACACGACATCGGGAAACAATGAGAGAGTTCGAGGGAGCCCTCTCTCAAC 154261
|||||
QY 121 CCTTGGCATTTCGGCAGTCTTCTGGGGCCGCTGAAGAGGCGGAGGCCAAGATGACCTT 180
|||||
Db 154260 CCTTGGCATTTCGGCAGTCTTCTGGGGCCGCTGAAGAGGCGGAGGCCAAGATGACCTT 154201
|||||
QY 181 CCAGATCTTTCAGCCCTTCTGAGCCAAAGAATTAGACGAAAGTGTCATTTGGCAAGA 240
|||||
Db 154200 CCAGATCTTTCAGCCCTTCTGAGCCAAAGAATTAGACGAAAGTGTCATTTGGCAAGA 154141
|||||
QY 241 CTGGCCATCAATTACGACCTTTTGGAGAGGCGAGATCAAACTCAAGCTAGAAAACGACTT 300
|||||
Db 154140 CTGGCCATCAATTACGACCTTTTGGAGAGGCGAGATCAAACTCAAGCTAGAAAACGACTT 154081
|||||
QY 301 TCTCCTGATCAGATGAACACTCACCTTAATTTAAGTTTTCAGGCTTAACCTTCTGCGAGAT 360
|||||
Db 154080 TCTCCTGATCAGATGAACACTCACCTTAATTTAAGTTTTCAGGCTTAACCTTCTGCGAGAT 154021
|||||
QY 361 AACCTCGAAGTCCACAGCTCTAAAGAAGCCGCCAGGAGGCAAAAGGCCACAGTAT 420
|||||
Db 154020 AACCTCGAAGTCCACAGCTCTAAAGAAGCCGCCAGGAGGCAAAAGGCCACAGTAT 153961
|||||
QY 421 TGTCTGAACCCAGTCCAAAAGTATTTTAAATAAGCTCCGACTTCATTGAAGAG 480
|||||
Db 153960 TGTCTGAACCCAGTCCAAAAGTATTTTAAATAAGCTCCGACTTCATTGAAGAG 153901
|||||
QY 481 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGTGGCTGCAAAAAC 540
|||||
Db 153900 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGTGGCTGCAAAAAC 153841
|||||
QY 541 CACAAGAGTAACTGGAAATCTCAAAACAAGTTATGACGAGAAAACAGCTCCAGTTTCTCA 600
|||||
Db 153840 CACAAGAGTAACTGGAAATCTCAAAACAAGTTATGACGAGAAAACAGCTCCAGTTTCTCA 153781
|||||
QY 601 GATCTGTCAAAAGACGAGAAAGTCTTCTGTCCCATCCCTATCCCTCGGATACACAGG 660
|||||
Db 153780 GATCTGTCAAAAGACGAGAAAGTCTTCTGTCCCATCCCTATCCCTCGGATACACAGG 153721
|||||
QY 661 GATAATGAAGTGAATCACGCCCTTGAACACGACGAGGAAGCCGCTGAAGCGGAGCAG 720
|||||
Db 153720 GATAATGAAGTGAATCACGCCCTTGAACACGACGAGGAAGCCGCTGAAGCGGAGCAG 153661
|||||
QY 721 GCTGCCAGTGAAGCGGCTGTGGAGACACTACACAGGGTCTTCCCTCTCATCTCTGTAC 780
|||||
Db 153660 GCTGCCAGTGAAGCGGCTGTGGAGACACTACACAGGGTCTTCCCTCTCATCTCTGTAC 153601
|||||
QY 781 TATCAAGAACCTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGAGCAGAGAGATT 840
|||||
Db 153600 TATCAAGAACCTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGAGCAGAGAGATT 153541
|||||
QY 841 CCAGAAGGAACTCGAGTACAGTTGGATTGCTAGGATTAGGAATTCACCAACCTCAAGTA 900
```

```
|||||
Db 153540 CCAGAAGCACTGAGTACAGTTGGTAGTGGTAGGAATTCACCACCTCAAGTA 153481
Qy 901 AAGT 904
|||||
Db 153480 AAGT 153477

RESULT 9
AL512429 11170 bp DNA linear PRI 26-OCT-2001
LOCUS Human DNA sequence from clone RP11-429C3 on chromosome 10, complete
DEFINITION sequence.
ACCESSION AL512429
VERSION AL512429.14 GI:16508239
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 11170)
AUTHORS Mashreghi-Mohammadi, M.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Oct 29, 2001 this sequence version replaced gi:15028699.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-429C3 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-429C3 It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-153K11 is at 9171 in this sequence.
The true right end of clone RP11-297N15 is at 2000 in this
sequence.

FEATURES
source Location/Qualifiers
1. .11170
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-429C3"
/clone_lib="RPCI-11.2"
1620. .2069
/note="match: GSS: Em:AQ435766"
complement(2076..2585)
/note="match: GSS: Em:AQ822926"
2596. .3437
/note="match: GSS: Em:AQ746971"
3298. .3591

misc_feature
/note="match: GSS: Em:AQ435766"
complement(2076..2585)
misc_feature
/note="match: GSS: Em:AQ822926"
2596. .3437
misc_feature
/note="match: GSS: Em:AQ746971"
3298. .3591
```

```
/note="match: GSS: Em:AQ742982"
3596. .3953
/note="match: GSS: Em:AQ742982"
complement(7993..8507)
/note="match: GSS: Em:AQ881403"
9218. .9756
/note="match: GSS: Em:AQ386388"
BASE COUNT 3230 a 2428 c 2215 g 3297 t
ORIGIN

Query Match 10.4%; Score 413; DB 9; Length 111170;
Best Local Similarity 99.5%; Pred. No. 5.9e-218;
Matches 563; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1974 TGATTCCACTCAGTTACAACAGCTTCATACCAAGTCTTACTGGAACAACACCAATGCA 2033
Db 4075 TGATTCCACTCAGTTACAACAGCTTCATACCAAGTCTTACTGGAACAACACCAATGCA 4134
Qy 2034 AAACCCACCTCCTTCATCTCTCTAAGAGGATTTCTCTTCARCATGACTGTTTGAACCTCAA 2093
Db 4135 AAACCCACCTCCTTCATCTCTCTAAGAGGATTTCTCTTCAGCATGACTGTTTGAACCTCAA 4194
Qy 2094 TGCTCCCGCCAGCGGTGACACATCCATAGCAGGTGAAGGTCCTTCATCACAGAGGTT 2153
Db 4195 TGCTCCCGCCAGCGGTGACACATCCATAGCAGGTGAAGGTCCTTCATCACAGAGGTT 4254
Qy 2154 CAGCTTGGCCGCGCGGAAGTATTTCTCCCTCCACCAACACCAACCGCAGCACTGTGGC 2213
Db 4255 CAGCTTGGCCGCGCGGAAGTATTTCTCCCTCCACCAACACCAACCGCAGCACTGTGGC 4314
Qy 2214 CCCTTCCAGCTCTCCGGTGTGTACTTTGAGCAGCACCTCTCAAACTATTTCAGAGGACAGT 2273
Db 4315 CCCTTCCAGCTCTCCGGTGTGTACTTTGAGCAGCACCTCTCAAACTATTTCAGAGGACAGT 4374
Qy 2274 GAGCAAGAAGGCTCTTAGTGTCTCACCCCTCTGTGCAAAACCAAAATCTCCAGAGGGCT 2333
Db 4375 GAGCAAGAAGGCTCTTAGTGTCTCACCCCTCTGTGCAAAACCAAAATCTCCAGAGGGCT 4434
Qy 2334 TTCCATCAAAATGAGCCACTCCACCAAGGCGCCCAAGACCAACCAACCACTTCAC 2393
Db 4435 TTCCATCAAAATGAGCCACTCCACCAAGGCGCCCAAGACCAACCAACCACTTCAC 4494
Qy 2394 ATTTTCCATCCCAAGCGGAACACAGTTTCAGCCCGCGTGTGTGTCCTCCCAATTCCTGCTC 2453
Db 4495 ATTTTCCATCCCAAGCGGAACACAGTTTCAGCCCGCGTGTGTGTCCTCCCAATTCCTGCTC 4554
Qy 2454 TCCTACCAAGCGGATTCAGAACCCAGTGGCTTTCCTCAGCTCTGTTGCTGCTCTCTCCC 2513
Db 4555 TCCTACCAAGCGGATTCAGAACCCAGTGGCTTTCCTCAGCTCTGTTGCTGCTCTCTCCC 4614
Qy 2514 TGCCATCCCAAGCGGAACACAGTTTCAGCCCGCGTGTGTGTCCTCCCAATTCCTGCTC 2539
Db 4615 TGCCATCCCAAGCGGAACACAGTTTCAGCCCGCGTGTGTGTCCTCCCAATTCCTGCTC 4640

RESULT 10
AL442068 290625 bp DNA linear HTG 20-NOV-2001
LOCUS Homo sapiens chromosome 10 clone RP11-331C8, *** SEQUENCING IN
DEFINITION PROGRESS ***, 43 unordered pieces.
ACCESSION AL442068
VERSION AL442068.16 GI:17043795
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Sims, S.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
```

COMMENT

On Nov 21, 2001 this sequence version replaced gi:14575284.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA331G8

----- Summary Statistics

Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 277809 bases at least Q40
Consensus quality: 281712 bases at least Q30
Consensus quality: 284299 bases at least Q20
Insert size: 286425; sum-of-contigs
Insert size: 156064; 7.4% error; agarose-fp
Quality coverage: 5.79x in Q20 bases; sum-of-contigs Quality
coverage: 15.47x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 103320: contig of 103320 bp in length
* 103321 103420: gap of 100 bp
* 103421 140724: contig of 37304 bp in length
* 140725 140824: gap of 100 bp
* 140825 162687: contig of 21863 bp in length
* 162688 162787: gap of 100 bp
* 162788 165269: contig of 2482 bp in length
* 165270 165369: gap of 100 bp
* 165370 167787: contig of 2418 bp in length
* 167788 167887: gap of 100 bp
* 167888 170672: contig of 2785 bp in length
* 170673 170772: gap of 100 bp
* 170773 173340: contig of 2568 bp in length
* 173341 173440: gap of 100 bp
* 173441 176369: contig of 2929 bp in length
* 176370 176469: gap of 100 bp
* 176470 179660: contig of 3191 bp in length
* 179661 179760: gap of 100 bp
* 179761 181788: contig of 2028 bp in length
* 181789 181888: gap of 100 bp
* 181889 185591: contig of 3703 bp in length
* 185592 185691: gap of 100 bp
* 185692 187897: contig of 2206 bp in length
* 187898 187997: gap of 100 bp
* 187998 190151: contig of 2134 bp in length
* 190152 190251: gap of 100 bp
* 190252 192977: contig of 2726 bp in length
* 192978 193077: gap of 100 bp
* 193078 195504: contig of 2427 bp in length
* 195505 195604: gap of 100 bp
* 195605 198738: contig of 3134 bp in length
* 198739 198838: gap of 100 bp
* 198839 202505: contig of 3667 bp in length
* 202506 202605: gap of 100 bp
* 202606 206757: contig of 4152 bp in length
* 206758 206857: gap of 100 bp
* 206858 211910: contig of 5053 bp in length
* 211911 212010: gap of 100 bp
* 212011 218152: contig of 6142 bp in length
* 218153 218252: gap of 100 bp
* 218253 221882: contig of 3630 bp in length
* 221883 221982: gap of 100 bp
* 221983 225735: contig of 3753 bp in length
* 225736 225835: gap of 100 bp
* 225836 229907: contig of 4072 bp in length

* 229908 230007: gap of 100 bp
* 230008 233016: contig of 3009 bp in length
* 233017 233116: gap of 100 bp
* 233117 236145: contig of 3029 bp in length
* 236146 236245: gap of 100 bp
* 236246 239609: contig of 3364 bp in length
* 239610 239709: gap of 100 bp
* 239710 241796: contig of 2087 bp in length
* 241797 241896: gap of 100 bp
* 241897 243915: contig of 2019 bp in length
* 243916 244015: gap of 100 bp
* 244016 246442: contig of 2427 bp in length
* 246443 246542: gap of 100 bp
* 246543 249599: contig of 3057 bp in length
* 249600 249699: gap of 100 bp
* 249700 252680: contig of 2981 bp in length
* 252681 252780: gap of 100 bp
* 252781 256312: contig of 3532 bp in length
* 256313 256412: gap of 100 bp
* 256413 259563: contig of 3151 bp in length
* 259564 259663: gap of 100 bp
* 259664 262204: contig of 2541 bp in length
* 262205 262304: gap of 100 bp
* 262305 264799: contig of 2495 bp in length
* 264800 264899: gap of 100 bp
* 264900 267645: contig of 2746 bp in length
* 267646 267745: gap of 100 bp
* 267746 271525: contig of 3780 bp in length
* 271526 271625: gap of 100 bp
* 271626 275798: contig of 4173 bp in length
* 275799 275898: gap of 100 bp
* 275899 278025: contig of 2127 bp in length
* 278026 278125: gap of 100 bp
* 278126 281433: contig of 3308 bp in length
* 281434 281533: gap of 100 bp
* 281534 284437: contig of 2904 bp in length
* 284438 284537: gap of 100 bp
* 284538 286727: contig of 2190 bp in length
* 286728 286827: gap of 100 bp
* 286828 290625: contig of 3798 bp in length.

FEATURES
 Location/Qualifiers
 1..290625
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-331G8"
 /clone_lib="RPCI-11.2"
 1..103320
 /note="assembly_fragment:07950
 fragment_chain:1
 clone_end:T7
 vector_side:left"
 103421..140724
 /note="assembly_fragment:01658
 fragment_chain:1"
 140825..162687
 /note="assembly_fragment:01866
 fragment_chain:1"
 162788..165269
 /note="assembly_fragment:01933
 fragment_chain:2"
 165370..167787
 /note="assembly_fragment:02386
 fragment_chain:2"
 167888..170672
 /note="assembly_fragment:02267
 fragment_chain:3"
 170773..173340
 /note="assembly_fragment:02094
 fragment_chain:3"
 173441..176369
 /note="assembly_fragment:00062"
 176470..179660

[illegible]

* 9513 10563: contig of 1051 bp in length
* 10564 10563: gap of unknown length
* 10664 11995: contig of 1332 bp in length
* 11996 12095: gap of unknown length
* 12096 13309: contig of 1214 bp in length
* 13310 13409: gap of unknown length
* 13410 15172: contig of 1763 bp in length
* 15173 15272: gap of unknown length
* 15273 17973: contig of 2701 bp in length
* 17974 18073: gap of unknown length
* 18074 21419: contig of 3346 bp in length
* 21420 21519: gap of unknown length
* 21520 24881: contig of 3362 bp in length
* 24882 24981: gap of unknown length
* 24982 31876: contig of 6895 bp in length
* 31877 31976: gap of unknown length
* 31977 49015: contig of 17039 bp in length
* 49016 49115: gap of unknown length
* 49116 66286: contig of 17171 bp in length
* 66287 66386: gap of unknown length
* 66387 86201: contig of 19815 bp in length
* 86202 86301: gap of unknown length
* 86302 132480: contig of 46179 bp in length
* 132481 132580: gap of unknown length
* 132581 175265: contig of 42685 bp in length.

FEATURES
source

1. .175265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-31303"
/clone_lib="RPC1-11"
1. .1145
/note="assembly_name:Contig1"
1246. .2414
/note="assembly_name:Contig11"
2515. .3864
/note="assembly_name:Contig12"
3965. .5078
/note="assembly_name:Contig15"
5179. .8647
/note="assembly_name:Contig18"
6748. .7926
/note="assembly_name:Contig20"
8027. .9412
/note="assembly_name:Contig23"
9513. .10563
/note="assembly_name:Contig25"
10664. .11995
/note="assembly_name:Contig27"
12096. .13309
/note="assembly_name:Contig28"
13410. .15172
/note="assembly_name:Contig29"
15273. .17973
/note="assembly_name:Contig30"
18074. .21419
/note="assembly_name:Contig31"
21520. .24881
/note="assembly_name:Contig32"
24982. .31876
/note="assembly_name:Contig33"
31977. .49015
/note="assembly_name:Contig34"
49116. .66286
/note="assembly_name:Contig35
clone_end:SP6"
66387. .86201
/note="assembly_name:Contig36"
86302. .132480
/note="assembly_name:Contig37
clone_end:77"
132581. .175265

/note="assembly_name:Contig38"
BASE COUNT 53132 a 34451 c 33650 g 52102 t 1930 others
ORIGIN
Query Match 6.0%; Score 236; DB 2; Length 175265;
Best Local Similarity 100.0%; Pred. No. 2.3e-119;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1648 ACCACCAACCTGGCAGCTATTGAGCCACAGCCCTCCACCCACCTCAGAGCCTCCATCT 1707
|||||
Db 54760 ACCACCAACCTGGCAGCTATTGAGCCACAGCCCTCCACCCACCTCAGAGCCTCCATCT 54701
|||||
QY 1708 GTGGAACAAACCCCAACCCAACTCGAGGGGGTCTTGTGTGAACCAACCAATGAGCCCGG 1767
|||||
Db 54700 GTGGAACAAACCCCAACCCAACTCGAGGGGGTCTTGTGTGAACCAACCAATGAGCCCGG 54641
|||||
QY 1768 TCCAGCTCCAGGATTCGGGCTTCGTGTGCACCTTCAACCTGCCTGAAGATGACAAGGAAGT 1827
|||||
Db 54640 TCCAGCTCCAGGATTCGGGCTTCGTGTGCACCTTCAACCTGCCTGAAGATGACAAGGAAGT 54581
|||||
QY 1828 GAAGCATCTCCGAGGCTGGTGTGGTGACCAACAGACAGACAGCCGCGGATTCCTTT 1883
|||||
Db 54580 GAAGCATCTCCGAGGCTGGTGTGGTGACCAACAGACAGACAGCCGCGGATTCCTTT 54525
|||||
RESULT 12
AL442068/c 290625 bp DNA linear HTG 20-NOV-2001
LOCUS
DEFINITION
Homo sapiens chromosome 10 clone RP11-331G8, *** SEQUENCING IN
PROGRESS ***, 43 unordered pieces.
ACCESSION AL442068
VERSION AL442068.16 GI:17043795
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Sims, S.
Direct Submission
Submitted (19-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 21, 2001 this sequence version replaced gi:14575284.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA331G8
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 277809 bases at least Q40
Consensus quality: 281712 bases at least Q30
Consensus quality: 284299 bases at least Q20
Insert size: 286425; sum-of-contigs
Quality coverage: 5.79x in Q20 bases; agarose-fp
coverage: 15.47x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 103320: contig of 103320 bp in length
* 103321 103420: gap of 100 bp

```

* 103421 140724: contig of 37304 bp in length
* 140725 140824: gap of 100 bp
* 140825 162687: contig of 21863 bp in length
* 162688 162787: gap of 100 bp
* 162788 165269: contig of 2482 bp in length
* 165270 165369: gap of 100 bp
* 165370 167787: contig of 2418 bp in length
* 167788 167887: gap of 100 bp
* 167888 170672: contig of 2785 bp in length
* 170673 170772: gap of 100 bp
* 170773 173340: contig of 2568 bp in length
* 173341 173440: gap of 100 bp
* 173441 176369: contig of 2929 bp in length
* 176370 176469: gap of 100 bp
* 176470 179660: contig of 3191 bp in length
* 179661 179760: gap of 100 bp
* 179761 181788: contig of 2028 bp in length
* 181789 181888: gap of 100 bp
* 181889 185591: contig of 3703 bp in length
* 185592 185691: gap of 100 bp
* 185692 187897: contig of 2206 bp in length
* 187898 187997: gap of 100 bp
* 187998 190151: contig of 2154 bp in length
* 190152 190251: gap of 100 bp
* 190252 192977: contig of 2726 bp in length
* 192978 193077: gap of 100 bp
* 193078 195504: contig of 2427 bp in length
* 195505 195604: gap of 100 bp
* 195605 198738: contig of 3134 bp in length
* 198739 198838: gap of 100 bp
* 198839 202505: contig of 3667 bp in length
* 202506 202605: gap of 100 bp
* 202606 206757: contig of 4152 bp in length
* 206758 206857: gap of 100 bp
* 206858 211910: contig of 5053 bp in length
* 211911 212010: gap of 100 bp
* 212011 218152: contig of 6142 bp in length
* 218153 218252: gap of 100 bp
* 218253 221882: contig of 3630 bp in length
* 221883 221982: gap of 100 bp
* 221983 225735: contig of 3753 bp in length
* 225736 225835: gap of 100 bp
* 225836 229907: contig of 4072 bp in length
* 229908 230007: gap of 100 bp
* 230008 233016: contig of 3009 bp in length
* 233017 233116: gap of 100 bp
* 233117 236145: contig of 3029 bp in length
* 236146 236245: gap of 100 bp
* 236246 239609: contig of 3364 bp in length
* 239610 239709: gap of 100 bp
* 239710 241796: contig of 2087 bp in length
* 241797 241896: gap of 100 bp
* 241897 243915: contig of 2019 bp in length
* 243916 244015: gap of 100 bp
* 244016 246442: contig of 2427 bp in length
* 246443 246542: gap of 100 bp
* 246543 249599: contig of 3057 bp in length
* 249600 249699: gap of 100 bp
* 249700 252680: contig of 2981 bp in length
* 252681 252780: gap of 100 bp
* 252781 256312: contig of 3532 bp in length
* 256313 256412: gap of 100 bp
* 256413 259563: contig of 3151 bp in length
* 259564 259663: gap of 100 bp
* 259664 262204: contig of 2541 bp in length
* 262205 262304: gap of 100 bp
* 262305 264799: contig of 2495 bp in length
* 264800 264899: gap of 100 bp
* 264900 267645: contig of 2746 bp in length
* 267646 267745: gap of 100 bp
* 267746 271525: contig of 3780 bp in length
* 271526 271625: gap of 100 bp
* 271626 275798: contig of 4173 bp in length

```

```

* 275799 275898: gap of 100 bp
* 275899 278025: contig of 2127 bp in length
* 278026 278125: gap of 100 bp
* 278126 281433: contig of 3308 bp in length
* 281434 281533: gap of 100 bp
* 281534 284437: contig of 2904 bp in length
* 284438 284537: gap of 100 bp
* 284538 286727: contig of 2190 bp in length
* 286728 286827: gap of 100 bp
* 286828 290625: contig of 3798 bp in length.
FEATURES
    Location/Qualifiers
        1..290625
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="RP11-331C8"
            /clone_lib="RPCI-11.2"
        1..103320
            /note="assembly_fragment:07950"
            fragment_chain:1
            clone_end:T7
            vector_side:left
            103421..140724
                /note="assembly_fragment:01658"
                fragment_chain:1
            140825..162687
                /note="assembly_fragment:01866"
                fragment_chain:1
            162788..165269
                /note="assembly_fragment:01933"
                fragment_chain:2
            165370..167787
                /note="assembly_fragment:02386"
                fragment_chain:2
            167888..170672
                /note="assembly_fragment:02267"
                fragment_chain:3
            170773..173340
                /note="assembly_fragment:02094"
                fragment_chain:3
            173441..176369
                /note="assembly_fragment:00062"
                fragment_chain:3
            176470..179660
                /note="assembly_fragment:00157"
                fragment_chain:3
            179761..181788
                /note="assembly_fragment:00471"
                fragment_chain:3
            181889..185591
                /note="assembly_fragment:00485"
                fragment_chain:3
            185692..187897
                /note="assembly_fragment:00607"
                fragment_chain:3
            187998..190151
                /note="assembly_fragment:01019"
                fragment_chain:3
            190252..192977
                /note="assembly_fragment:01432"
                fragment_chain:3
            193078..195504
                /note="assembly_fragment:01879"
                fragment_chain:3
            195605..198738
                /note="assembly_fragment:01944"
                fragment_chain:3
            198839..202505
                /note="assembly_fragment:02174"
                fragment_chain:3
            202606..206757
                /note="assembly_fragment:02200"
                fragment_chain:3
            206858..211910
                /note="assembly_fragment:02684"
                fragment_chain:3
            212011..218152
                /note="assembly_fragment:03379"
                fragment_chain:3
            218253..221882
                /note="assembly_fragment:03527"
                fragment_chain:3
            221983..225735
                /note="assembly_fragment:03880"
                fragment_chain:3
            225836..229907
                /note="assembly_fragment:04098"
                fragment_chain:3
            230008..233016

```

misc_feature /note="assembly_fragment:04166"
233117..236145
misc_feature /note="assembly_fragment:04570"
236246..239603
misc_feature /note="assembly_fragment:04709"
239710..241796
misc_feature /note="assembly_fragment:04939"
241897..243915

Query Match 5.7%; Score 227; DB 2; Length 290625;
Best Local Similarity 100.0%; Pred. No. 2.2e-114;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1648 ACCACCAACTGCGACCTATTGAGCCACAGCCCTCCACCCACCTCAGAGCGCTCCATCT 1707
Db 285418 ACCACCAACTGCGACCTATTGAGCCACAGCCCTCCACCCACCTCAGAGCGCTCCATCT 285359
QY 1708 GTGGAAACACCCCCCAACCCAAACCTCGAGGGGGTTCCTGGTGAACACCAATGAGCCCCGG 1767
Db 285358 GTGGAAACACCCCCCAACCCAAACCTCGAGGGGGTTCCTGGTGAACACCAATGAGCCCCGG 285299
QY 1768 TTCAGCTCCAGGATTCGGCTTCGTGTCACCTCAACCTGCCTGAAGATGACAAAGCAAT 1827
Db 285298 TTCAGCTCCAGGATTCGGCTTCGTGTCACCTCAACCTGCCTGAAGATGACAAAGCAAT 285239
QY 1828 GAAGCATCTCCGAGGCTGGTGTGGTGACCCAGACAGACACGAGGCC 1874
Db 285238 GAAGCATCTCCGAGGCTGGTGTGGTGACCCAGACAGACACGAGGCC 285192

RESULT 13
AC016395/c
LOCUS AC016395 166869 bp DNA linear PRI 14-MAY-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-153K11, complete sequence.
VERSION AC016395.5 GI:20564406
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166869)
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166869)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 166869)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 166869)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 14, 2002 this sequence version replaced gi:9929646.
FEATURES
source 1..166869
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-153K11"
/clone_lib="RPCL-11"

BASE COUNT 47408 a 35216 c 35336 g 48909 t
ORIGIN

Query Match 5.3%; Score 210; DB 9; Length 166869;

Best Local Similarity 100.0%; Pred. No. 7.1e-105;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3284 AGGTGAGTGGTTTACCGCCCGGAGCTGACATGGCTACTCAATGCGCAACCTGTCTCTAC 3343
Db 146665 AGGTGAGTGGTTTACCGCCCGGAGCTGACATGGCTACTCAATGCGCAACCTGTCTCTAC 146606
QY 3344 CAGATGCCTCCCAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTGACC 3403
Db 146605 CAGATGCCTCCCAAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTGACC 146546
QY 3404 CACTCACTCAGCGCAGCAGGAGACCTATAAGTGCATCGCTACCAACAAACCGGCAGA 3463
Db 146545 CACTCACTCAGCGCAGCAGGAGACCTATAAGTGCATCGCTACCAACAAACCGGCAGA 146486
QY 3464 ATTCTTTTGTAGCTGGAGCTCTCTGTAGTAG 3493
Db 146485 ATTCTTTTGTAGCTGGAGCTCTCTGTAGTAG 146456

RESULT 14
AC114220/c
LOCUS AC114220 151002 bp DNA linear HTG 17-JUL-2002
DEFINITION Rattus norvegicus clone CH230-318023, *** SEQUENCING IN PROGRESS
***, 43 unordered pieces.
AC114220
VERSION AC114220.2 GI:21744220
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 151002)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,F.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.E.,
Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,B., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooftari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

LOCUS AC114220 151002 bp DNA linear HTG 17-JUL-2002
DEFINITION Rattus norvegicus clone CH230-318023, *** SEQUENCING IN PROGRESS
AC114220 *** 43 unordered pieces.
VERSION AC114220.2 GI:21744220
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 151002)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Ayelle,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denna,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holliday,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozador,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,P., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 151002)
Worley,K.C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 151002)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:19224547.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GQWH
Center clone name: CH230-318023
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 109622 bases at least Q40
Consensus quality: 113752 bases at least Q30
Consensus quality: 117081 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1050: contig of 1050 bp in length
* 1051 1150: gap of unknown length
* 1151 2234: contig of 1084 bp in length
* 2235 2334: gap of unknown length
* 2335 3615: contig of 1181 bp in length
* 3616 4928: contig of 1313 bp in length
* 4929 5028: gap of unknown length
* 5029 6455: contig of 1427 bp in length
* 6456 6555: gap of unknown length
* 6556 7769: contig of 1213 bp in length
* 7769 8880: gap of unknown length
* 8880 8880: contig of 1012 bp in length
* 8881 8980: gap of unknown length
* 8981 10397: contig of 1416 bp in length
* 10397 10496: gap of unknown length
* 10497 12153: contig of 1657 bp in length
* 12154 12253: gap of unknown length
* 12254 13855: contig of 1602 bp in length
* 13856 13955: gap of unknown length
* 13956 15469: contig of 1514 bp in length
* 15470 15569: gap of unknown length
* 15570 17115: contig of 1546 bp in length
* 17116 17215: gap of unknown length
* 17216 18930: contig of 1715 bp in length
* 18931 19030: gap of unknown length
* 19031 20071: contig of 1041 bp in length
* 20072 20171: gap of unknown length
* 20172 21498: contig of 1327 bp in length
* 21499 21598: gap of unknown length
* 21599 23109: contig of 1411 bp in length
* 23110 23109: gap of unknown length
* 23110 25272: contig of 2163 bp in length
* 25273 25372: gap of unknown length
* 25373 27230: contig of 1858 bp in length
* 27231 27330: gap of unknown length
* 27331 28961: contig of 1631 bp in length
* 28962 29061: gap of unknown length
* 29062 31889: contig of 2828 bp in length
* 31890 31889: gap of unknown length
* 31890 34607: contig of 2618 bp in length
* 34608 34707: gap of unknown length
* 34708 37331: contig of 2624 bp in length
* 37332 37431: gap of unknown length
* 37432 39898: contig of 2467 bp in length
* 39899 39998: gap of unknown length
* 39999 42294: contig of 2296 bp in length
* 42295 42394: gap of unknown length
* 42395 45156: contig of 2762 bp in length
* 45157 45256: gap of unknown length
* 45257 48350: contig of 3094 bp in length
* 48351 48450: gap of unknown length
* 48451 52367: contig of 3917 bp in length

* 52368 52467: gap of unknown length
* 52468 54222: contig of 1755 bp in length
* 54323 54322: gap of unknown length
* 54323 57293: contig of 2971 bp in length
* 57294 57393: gap of unknown length
* 57394 60595: contig of 3202 bp in length
* 60596 60695: gap of unknown length
* 60696 64285: contig of 3590 bp in length
* 64286 64385: gap of unknown length
* 64386 68430: contig of 4045 bp in length
* 68431 68530: gap of unknown length
* 68531 72515: contig of 3985 bp in length
* 72516 72616: gap of unknown length
* 72616 78326: contig of 5711 bp in length
* 78327 78426: gap of unknown length
* 78427 83087: contig of 4661 bp in length
* 83088 83187: gap of unknown length
* 83188 89628: contig of 6441 bp in length
* 89629 89728: gap of unknown length
* 89729 96246: contig of 6518 bp in length
* 96247 96346: gap of unknown length
* 96347 103840: contig of 7494 bp in length
* 103841 103940: gap of unknown length
* 103941 112875: contig of 8935 bp in length
* 112876 112975: gap of unknown length
* 112976 121627: contig of 8652 bp in length
* 121628 121727: gap of unknown length
* 121728 131119: contig of 9392 bp in length
* 131120 131219: gap of unknown length
* 131220 139984: contig of 8765 bp in length
* 139985 140084: gap of unknown length
* 140085 151002: contig of 10918 bp in length.

FEATURES Location/Qualifiers
source 1..151002
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-318023"

BASE COUNT 41237 a 32621 c 32087 g 40808 t 4249 others
ORIGIN

Query Match 1.0%; Score 38; DB 2; Length 151002;
Best Local Similarity 100.0%; Pred. No. 5.9e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2890 TTCACCTGCAAAATTGTTGGGATACCTGTTCCAAAGGT 2927
|||||
Db 41465 TTCACCTGCAAAATTGTTGGGATACCTGTTCCAAAGGT 41502

Search completed: November 30, 2002, 16:55:12
Job time : 9637 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 13:41:14 ; Search time 516 Seconds
(without alignments)
17295.871 Million cell updates/sec

Title: US-09-818-990B-1
Perfect score: 3963
Sequence: 1 atgcaagacgacagcataga.....tggagagtgatgaactttaa 3963

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :				N_Geneseq_l01002:*	
1:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*				
2:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*				
3:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*				
4:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*				
5:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*				
6:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*				
7:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*				
8:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*				
9:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*				
10:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*				
11:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*				
12:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*				
13:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*				
14:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*				
15:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*				
16:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*				
17:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*				
18:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*				
19:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*				
20:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*				
21:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*				
22:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*				
23:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*				
24:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	ID		
1	1483	37.4	2127	22	AAH13889	Human cDNA sequence
2	1056	26.6	1228	23	AAH74798	DNA encoding novel
3	787	19.9	2059	22	ABA06552	Human cDNA SEQ ID
4	685	17.3	2232	22	AAH28777	Human immunoglobulin
5	438	11.1	855	22	AAH04212	Human cDNA clone (
6	348	8.8	702	22	ABA06731	Human cDNA SEQ ID
7	348	8.8	702	22	AAH28841	Human immunoglobulin
8	318	8.0	392	21	AAH75084	Human ORF ORF639
9	99	2.5	200	24	ABN25334	Human ORF polynuc

530	22	AAH09723	Human cDNA clone (
281	24	ABL75308	Corn tassal-derive
300	20	AAZ13687	Human gene express
399	20	AAZ09013	Differentiation in
551	24	ABQ57885	Human colon cancer
733	20	AAZ15038	Human gene express
775	22	AAI96358	Human neuroblastom
1399	12	AAQ11843	Human lactoperoxid
2732	22	AAH33343	Human colon cancer
102	15	AAQ76487	Human genome fragm
324	22	AAI88440	Human polynucleoti
360	20	AAH84163	DNA encoding human
360	21	AAH79392	5' cDNA sequence o
360	24	ABK28936	Human breast tumou
486	20	AAH39605	Breast cancer asso
487	22	AAH14400	Human breast cancer
531	22	AAH14497	Human breast cancer
626	23	AAH76943	DNA encoding novel
653	22	ABA08994	Human integral mem
694	22	AAH14423	Human breast cancer
743	22	AAH23262	Human breast cancer
747	20	AAH40027	Prostate cancer as
755	22	AAH23285	Human breast cancer
771	22	AAH23359	Human breast cancer
814	22	AAH06054	Human cDNA clone (
956	21	AAH39796	Human cDNA clone (
957	22	AAH98167	Arabidopsis thalia
1047	20	ABA96501	Rat EST-derived co
1228	24	ABK72953	Human extracellular
1296	22	AAH68454	Bacillus lichenifo
1419	22	AAH71655	Corynebacterium q1
1467	22	AAH71654	Corynebacterium q1
1608	24	ABL53686	Human pterin molyb
1611	19	AAH62479	Human MAP kinase,
1611	24	ABL59534	Human mitogen acti
1635	24	ABK34558	Human cDNA for nov
1767	23	ABK24452	Human prostate exp
1767	23	ABK28267	Human prostate exp
1823	21	ABH98221	Human colon cancer
1846	24	ABK88194	DNA encoding human
1977	22	AAH58523	HOAT4 DNA. Homo s
1986	22	AAH09561	Human transporter
2013	22	AAH34803	Human colon cancer
2027	21	AAH61892	cDNA encoding a hu
2135	14	AAQ37469	C-glutamicum isoci
2162	22	AAH17716	Human cDNA sequenc
2508	22	AAH57454	Human liver cell s
2791	24	ABK84759	Human cDNA differe
2922	22	AAH02397	Human secreted pro
3303	21	AAH81710	Human secreted pro
3447	21	AAH54160	Pseudomonas aerugi
4125	21	AAH76971	Human ORF ORF2526
4195	23	AAH59189	DNA encoding novel
5279	21	AAH59189	Human secreted pro
5724	22	ABA21347	Human nervous syst
5724	22	ABA21347	Human nervous syst
6314	14	AAH04341	Sequence of a cDNA
6314	24	ABL65009	Lung cancer relate
6314	24	ABL65920	Lung cancer relate
7314	23	ABL19196	Drosophila melanog
7316	23	ABL29930	Drosophila melanog
9990	22	AAH26123	Mouse prion protei
9990	22	AAH26123	Mouse PrP/human ta
10322	22	AAH27811	DNA encoding novel
349980	22	AAH68532	C glutamicum codin
60	24	ABN39429	Human spliced tran
178	22	ABA72756	Human foetal liver
178	22	ABA38402	Probe #16868 for g
178	22	AAH21186	Human brain expres
178	22	AAH47342	Human bone marrow
178	22	AAH53179	Probe #21865 used
178	24	ABH21499	Human genome-deriv
222	24	ABK97878	Mouse ARP-2 cDNA v

c 83	18	0.5	230	22	ABA70400	Human foetal liver	c 156	18	0.5	1241	22	AAK65784	Human immune/haema
c 84	18	0.5	230	22	AAK37061	Probe #1527 for g	c 157	18	0.5	1280	22	AAK54226	CDNA encoding nove
c 85	18	0.5	230	22	AAK18639	Human brain expres	c 158	18	0.5	1287	23	ABL16985	Drosophila melanog
c 86	18	0.5	230	22	AAK44566	Human bone marrow	c 159	18	0.5	1333	20	AAK00712	Human secreted pro
c 87	18	0.5	230	22	AAK124920	Probe #14853 for g	c 160	18	0.5	1342	21	AAK38803	Arabidopsis thalia
c 88	18	0.5	230	22	AAI50550	Probe #19236 used	c 161	18	0.5	1343	21	AAK49751	Arabidopsis thalia
c 89	18	0.5	230	22	ABS18803	Human genome-deriv	c 162	18	0.5	1352	21	AAK75885	Human ORFX ORF1440
c 90	18	0.5	291	23	ABK42781	Genomic sequence #	c 163	18	0.5	1671	22	AAH97737	Human protein enco
c 91	18	0.5	291	23	ABK42782	Genomic sequence #	c 164	18	0.5	1694	19	AAV33190	Secreted protein B
c 92	18	0.5	300	20	AAK13140	Human gene express	c 165	18	0.5	1694	19	AAV33190	Positive strand se
c 93	18	0.5	334	22	AAI80060	Human polynucleoti	c 166	18	0.5	1763	21	AAK26452	Human secreted pro
c 94	18	0.5	337	14	AAK61176	Human brain expres	c 167	18	0.5	1772	20	AAZ00457	Human secreted pro
c 95	18	0.5	392	22	AAK66821	Novel human polynu	c 168	18	0.5	1778	21	AAA37056	Human PRO1311 (UNQ
c 96	18	0.5	439	23	AAK58724	CDNA #1400 encodin	c 169	18	0.5	1778	22	AAK54063	Human DNA encoding
c 97	18	0.5	461	24	ABL80122	Human ovarian canc	c 170	18	0.5	1778	22	AAK54063	DNA encoding prote
c 98	18	0.5	476	22	ABA57784	Human foetal liver	c 171	18	0.5	1812	24	ABL90612	Human polynucleoti
c 99	18	0.5	476	22	ABA27146	Probe #5612 for ge	c 172	18	0.5	1815	20	AAK52238	Human polynucleoti
c 100	18	0.5	476	22	AAK05851	Human brain expres	c 173	18	0.5	1815	22	AAK52238	Human DNA encoding
c 101	18	0.5	476	22	AAK31484	Human bone marrow	c 174	18	0.5	1815	22	AAK52238	Human PRO234 CDN
c 102	18	0.5	476	22	AAI15696	Probe #5629 for ge	c 175	18	0.5	1839	22	AAK72396	Human PRO234 CDNA
c 103	18	0.5	476	22	AAI37368	Probe #6054 used t	c 176	18	0.5	1839	22	AAK72396	Vigna unguiculata
c 104	18	0.5	476	22	ABS06238	Human genome-deriv	c 177	18	0.5	2005	21	AAK36013	Human musculoskele
c 105	18	0.5	492	24	ABL82831	Human ovarian canc	c 178	18	0.5	2005	21	AAK36013	CDNA encoding a le
c 106	18	0.5	501	24	ABK75105	Bacillus lichenifo	c 179	18	0.5	2052	21	AAK297124	Human secreted pro
c 107	18	0.5	514	24	ABK73881	Bovine embryonic g	c 180	18	0.5	2124	22	AAH60481	Human polynucleoti
c 108	18	0.5	527	24	ABQ30936	Oligonucleotide fo	c 181	18	0.5	2154	22	AAH67126	C glutamicum codin
c 109	18	0.5	527	24	ABQ30937	Oligonucleotide fo	c 182	18	0.5	2169	23	ABL16391	Drosophila melanog
c 110	18	0.5	567	22	ABA63873	Human foetal liver	c 183	18	0.5	2172	23	AAK592579	DNA encoding novel
c 111	18	0.5	567	22	AAK38099	Human bone marrow	c 184	18	0.5	2179	23	ABL14901	Drosophila melanog
c 112	18	0.5	567	22	AAI18873	Probe #8806 for ge	c 185	18	0.5	2290	21	AAK93371	Human secreted pro
c 113	18	0.5	567	22	AAI43996	Probe #12682 used	c 186	18	0.5	2352	23	ABL05135	Drosophila melanog
c 114	18	0.5	579	18	AAV24787	H. pylori ORF 07cp	c 187	18	0.5	2387	23	AAK51862	Staphylococcus aur
c 115	18	0.5	579	24	AAI70674	Helicobacter pylor	c 188	18	0.5	2436	23	AAK574233	DNA encoding novel
c 116	18	0.5	585	21	AAK69696	Novel Human protei	c 189	18	0.5	2658	21	AAK69801	Human breast tumou
c 117	18	0.5	589	22	AAK11875	Human brain expres	c 190	18	0.5	2715	23	AAK54672	Staphylococcus aur
c 118	18	0.5	593	22	ABA60217	Human foetal liver	c 191	18	0.5	2750	22	AAK02607	Pseudomonas exotox
c 119	18	0.5	593	22	ABA28525	Probe #6991 for ge	c 192	18	0.5	2760	17	AAV28494	P. aeruginosa dete
c 120	18	0.5	593	22	AAK08494	Human brain expres	c 193	18	0.5	2760	19	AAV61784	Pseudomonas aerugi
c 121	18	0.5	593	22	AAK34376	Human bone marrow	c 194	18	0.5	3109	23	ABL18836	Pseudomonas aerugi
c 122	18	0.5	593	22	AAI40100	Probe #8786 used t	c 195	18	0.5	3157	21	AAK61881	Drosophila melanog
c 123	18	0.5	593	24	ABS09067	Human genome-deriv	c 196	18	0.5	3157	21	AAK61881	CDNA encoding huma
c 124	18	0.5	624	23	ABK41658	CDNA encoding nove	c 197	18	0.5	3190	20	AAV72213	Human NT2-1A DNA
c 125	18	0.5	662	23	AAH04105	Drosophila melanog	c 198	18	0.5	3265	24	ABL55333	Human protein phos
c 126	18	0.5	708	22	AAH06315	Human cDNA clone (c 199	18	0.5	3265	24	ABL56802	Protein phosphatas
c 127	18	0.5	711	11	AAQ06638	Coat protein gene	c 200	18	0.5	3440	23	ABL03072	Drosophila melanog
c 128	18	0.5	711	14	AAQ36839	Potato virus X coa	c 201	18	0.5	3451	17	AAK28492	P. aeruginosa dete
c 129	18	0.5	712	24	ABK50062	DNA encoding HIV-1	c 202	18	0.5	3451	22	ABA76842	Pseudomonas aerugi
c 130	18	0.5	769	20	AAI15221	Human gene express	c 203	18	0.5	3723	24	ABL92667	Chlamydia pneumoni
c 131	18	0.5	769	20	AAI15204	Human gene express	c 204	18	0.5	3734	23	ABL16986	Drosophila melanog
c 132	18	0.5	774	19	AAV16113	Synthetic BRSV Gsy	c 205	18	0.5	3854	23	ABL16984	Drosophila melanog
c 133	18	0.5	795	20	AAK32984	Human warts gene f	c 206	18	0.5	3865	19	AAV59789	Human secreted pro
c 134	18	0.5	835	23	ABV20208	Human prostate exp	c 207	18	0.5	3881	24	ABL61961	Colon adenocarcino
c 135	18	0.5	835	23	ABV26037	Human prostate exp	c 208	18	0.5	3881	24	ABL63705	Breast cancer rela
c 136	18	0.5	838	13	AAQ21150	Sequence encoding	c 209	18	0.5	3881	24	ABL64104	Breast cancer rela
c 137	18	0.5	840	22	AAK88495	Bovine RSV G-prote	c 210	18	0.5	3881	24	ABL68087	Ovary cancer relat
c 138	18	0.5	861	15	AAK68661	Human phosphatase	c 211	18	0.5	3913	22	AAI59784	Human polynucleoti
c 139	18	0.5	891	23	ABL18837	Drosophila melanog	c 212	18	0.5	3984	17	AAV42118	H-lats gene encodi
c 140	18	0.5	900	21	AAK65502	Porcine BAC-pigf2-	c 213	18	0.5	3984	21	AAK51505	Human Lats (large
c 141	18	0.5	903	24	ABN68341	Streptococcus poly	c 214	18	0.5	4071	22	AAI57998	Human polynucleoti
c 142	18	0.5	906	24	AAK31867	Lactobacillus rham	c 215	18	0.5	4159	22	AAK594746	Human DNA sequence
c 143	18	0.5	913	22	AAK40921	CDNA encoding nove	c 216	18	0.5	4175	22	AAH98543	Human EST-derived
c 144	18	0.5	999	21	AAK49304	Arabidopsis thalia	c 217	18	0.5	4191	19	AAV23692	L.lactis MG1363 pf
c 145	18	0.5	1045	22	AAK71247	Corynebacterium gl	c 218	18	0.5	4198	22	AAH98630	Human EST-derived
c 146	18	0.5	1066	22	AAK71246	Corynebacterium gl	c 219	18	0.5	4356	23	ABL06508	Drosophila melanog
c 147	18	0.5	1066	22	AAK71785	Corynebacterium gl	c 220	18	0.5	4373	23	ABL14900	Drosophila melanog
c 148	18	0.5	1066	23	AAK596104	C. glutamicum gene	c 221	18	0.5	4392	24	AAK62082	Human cDNA encodin
c 149	18	0.5	1117	22	AAH14400	Human cDNA sequenc	c 222	18	0.5	4398	21	AAK3063	Arabidopsis thalia
c 150	18	0.5	1186	21	AAA49461	Coat protein codin	c 223	18	0.5	4417	20	AAV72214	Human NT2-2A DNA
c 151	18	0.5	1191	20	AAK00660	Human secreted pro	c 224	18	0.5	4441	23	ABL16388	Drosophila melanog
c 152	18	0.5	1191	23	AAK591730	DNA encoding novel	c 225	18	0.5	4471	21	AAK16390	Drosophila melanog
c 153	18	0.5	1195	21	AAK40915	Arabidopsis thalia	c 226	18	0.5	4538	21	AAK77408	Human ORFX ORF2963
c 154	18	0.5	1205	21	AAK42243	Arabidopsis thalia	c 227	18	0.5	4538	22	AAK65785	Human immune/haema
c 155	18	0.5	1233	23	AAK69109	DNA encoding novel	c 228	18	0.5	4538	22	AAK65786	Human immune/haema

c 229	18	0.5	4580	21	AAc77082	Human ORFX ORF2637	302	17	0.4	250	20	AAx37718	Human clone 234541
c 230	18	0.5	4580	21	AAc77082	Human ORFX ORF2637	303	17	0.4	250	20	AAx56521	Human 2345419 DNA
c 231	18	0.5	4843	20	AAx37553	Human HLIG-1 cDNA.	304	17	0.4	252	20	AAx56525	Human 1932979 DNA
c 232	18	0.5	4890	23	AAx68250	DNA encoding novel	c 305	17	0.4	255	21	AAz58130	Human lactoferrin
c 233	18	0.5	5113	23	ABLO8880	Drosophila melanog	c 306	17	0.4	255	21	AAz58130	Human lactoferrin
c 234	18	0.5	5202	22	AAH99637	Human protein enco	c 307	17	0.4	261	20	AAx56522	Human 1731885 DNA
c 235	18	0.5	5671	22	AAz20915	Human nervous syst	c 308	17	0.4	268	22	AAx56522	Human 1731885 DNA
c 236	18	0.5	5671	22	AAx39832	Genomic sequence #	c 309	17	0.4	287	22	AAx38254	Novel human diago
c 237	18	0.5	5671	22	AAx39832	Human digestive sy	c 310	17	0.4	287	22	AAx38254	Novel human diago
c 238	18	0.5	6279	23	ABLO9108	Human digestive sy	c 311	17	0.4	291	13	AAQ22192	Human immune/haema
c 239	18	0.5	6279	23	ABLO9108	Drosophila melanog	c 312	17	0.4	291	13	AAQ32787	Tetranucleotide re
c 240	18	0.5	6927	22	ABAL6073	Human nervous syst	c 313	17	0.4	291	15	AAQ57810	Microsatellite rep
c 241	18	0.5	6930	22	ABAL6072	Human nervous syst	c 314	17	0.4	291	15	AAQ57810	Repeat polymorphis
c 242	18	0.5	6930	22	ABAL6072	Human nervous syst	c 315	17	0.4	291	15	AAQ57810	Polymorphic region
c 243	18	0.5	7382	20	AAx87396	Human nervous syst	c 316	17	0.4	291	24	ABL72180	Corn tassell-derive
c 244	18	0.5	7869	23	ABLO5134	Human WART1 cDNA.	c 317	17	0.4	291	24	ABL72180	Human ovarian canc
c 245	18	0.5	8148	22	ABAL9444	Drosophila melanog	c 318	17	0.4	293	20	AAx37719	Human clone T87045
c 246	18	0.5	8148	22	ABAL9444	Human nervous syst	c 319	17	0.4	294	23	ABV61699	Human prostate exp
c 247	18	0.5	8148	22	ABAL9444	Human nervous syst	c 320	17	0.4	310	22	ABA50745	Human breast cell
c 248	18	0.5	8188	24	ABK15597	Human reproductive	c 321	17	0.4	310	22	ABA50745	Human foetal liver
c 249	18	0.5	8188	24	ABK15597	Human testicular a	c 322	17	0.4	310	22	ABA50745	Human foetal liver
c 250	18	0.5	8188	24	ABK15597	DNA encoding novel	c 323	17	0.4	310	22	ABA50745	Human brain expres
c 251	18	0.5	8188	24	ABK15597	DNA encoding novel	c 324	17	0.4	310	22	ABA50745	Human bone marrow
c 252	18	0.5	8188	24	ABK15597	DNA encoding novel	c 325	17	0.4	310	22	ABA50745	Probe #13535 for g
c 253	18	0.5	8486	24	ABN83687	DNA encoding novel	c 326	17	0.4	310	22	ABA50745	Probe #17600 used
c 254	18	0.5	9100	24	ABN83688	Vector plasmid pCa	c 327	17	0.4	310	22	ABA50745	Probe #9208 used t
c 255	18	0.5	10118	22	AAx46252	DNA encoding novel	c 328	17	0.4	315	22	AAK21634	Human genome-deriv
c 256	18	0.5	10616	22	AAK85565	DNA encoding novel	c 329	17	0.4	315	22	AAK21634	Human foetal liver
c 257	18	0.5	10616	22	AAK85565	DNA encoding novel	c 330	17	0.4	315	22	AAK21634	Human brain expres
c 258	18	0.5	11206	21	AAz47807	Human immune/haema	c 331	17	0.4	315	22	AAK21634	Human bone marrow
c 259	18	0.5	11572	23	ABL11264	Drosophila melanog	c 332	17	0.4	315	22	AAK21634	Probe #13535 for g
c 260	18	0.5	11871	20	AAx13108	Enterococcus faeca	c 333	17	0.4	315	22	AAK21634	Probe #17600 used
c 261	18	0.5	14070	22	AAK70465	Human immune/haema	c 334	17	0.4	315	22	AAK21634	Probe #9208 used t
c 262	18	0.5	14283	23	ABL11798	Drosophila melanog	c 335	17	0.4	315	22	AAK21634	Human genome-deriv
c 263	18	0.5	16091	23	AAx89978	DNA encoding novel	c 336	17	0.4	315	22	AAK21634	Human foetal liver
c 264	18	0.5	17357	22	AAK86680	Human immune/haema	c 337	17	0.4	315	22	AAK21634	Human brain expres
c 265	18	0.5	17363	22	AAK86681	Human immune/haema	c 338	17	0.4	315	22	AAK21634	Human bone marrow
c 266	18	0.5	28149	22	ABAL7961	Human nervous syst	c 339	17	0.4	315	22	AAK21634	Probe #13535 for g
c 267	18	0.5	32192	22	ABAL7963	Human nervous syst	c 340	17	0.4	315	22	AAK21634	Probe #17600 used
c 268	18	0.5	66566	21	AAx53450	Human thioresoxin	c 341	17	0.4	315	22	AAK21634	Probe #9208 used t
c 269	18	0.5	112190	22	AAH44801	Human GPCR protein	c 342	17	0.4	315	22	AAK21634	Human genome-deriv
c 270	18	0.5	136284	24	ABK83575	Human cDNA differe	c 343	17	0.4	315	22	AAK21634	Human immune/haema
c 271	18	0.5	147419	24	ABK83574	Human cDNA differe	c 344	17	0.4	315	22	AAK21634	Human genome-deriv
c 272	18	0.5	349980	22	AAH41225	Pyrococcus abyssi	c 345	17	0.4	315	22	AAK21634	Human prostate exp
c 273	18	0.5	349980	22	AAH41225	Pyrococcus abyssi	c 346	17	0.4	315	22	AAK21634	Human prostate exp
c 274	18	0.5	611590	21	AAH22303	C glutamicum codin	c 347	17	0.4	315	22	AAK21634	Human secreted pro
c 275	17	0.4	20	21	AAz295339	Arabidopsis thalia	c 348	17	0.4	315	22	AAK21634	Human nervous syst
c 276	17	0.4	60	24	ABN38137	Human mtPEPCK phos	c 349	17	0.4	315	22	AAK21634	Human polynucleoti
c 277	17	0.4	60	24	ABN38137	Human mtPEPCK phos	c 350	17	0.4	315	22	AAK21634	Human foetal liver
c 278	17	0.4	142	22	ABA36502	Human spliced tran	c 351	17	0.4	315	22	AAK21634	Probe #5306 for ge
c 279	17	0.4	142	22	ABA36502	Human spliced tran	c 352	17	0.4	315	22	AAK21634	Human brain expres
c 280	17	0.4	142	22	ABA36502	Probe #14968 for g	c 353	17	0.4	315	22	AAK21634	Human bone marrow
c 281	17	0.4	149	21	AAc29324	Probe #14358 for g	c 354	17	0.4	315	22	AAK21634	Human bone marrow
c 282	17	0.4	153	22	ABAL1539	Human secreted pro	c 355	17	0.4	315	22	AAK21634	Probe #5557 for ge
c 283	17	0.4	153	24	ABAL25380	Human nervous syst	c 356	17	0.4	315	22	AAK21634	Probe #5884 used t
c 284	17	0.4	163	22	AAK18844	Human ORFX polynuc	c 357	17	0.4	315	22	AAK21634	Human genome-deriv
c 285	17	0.4	163	22	AAK18844	Human foetal liver	c 358	17	0.4	315	22	AAK21634	DNA encoding novel
c 286	17	0.4	163	22	AAK44786	Human brain expres	c 359	17	0.4	315	22	AAK21634	DNA encoding novel
c 287	17	0.4	163	22	AAK50763	Human bone marrow	c 360	17	0.4	315	22	AAK21634	Eucalyptus grandis
c 288	17	0.4	163	24	ABN19025	Probe #19449 used	c 361	17	0.4	315	22	AAK21634	Human breast cell
c 289	17	0.4	167	22	ABAL1172	Human genome-deriv	c 362	17	0.4	315	22	AAK21634	Human foetal liver
c 290	17	0.4	167	22	AAK25300	Probe #19638 for g	c 363	17	0.4	315	22	AAK21634	Probe #4245 for ge
c 291	17	0.4	167	22	AAK51307	Human brain expres	c 364	17	0.4	315	22	AAK21634	Human bone marrow
c 292	17	0.4	167	22	AAK128318	Human bone marrow	c 365	17	0.4	315	22	AAK21634	Human bone marrow
c 293	17	0.4	167	24	ABN24850	Probe #18251 for g	c 366	17	0.4	315	22	AAK21634	Probe #4334 for ge
c 294	17	0.4	175	24	ABN18320	Human genome-deriv	c 367	17	0.4	315	22	AAK21634	Probe #4460 used t
c 295	17	0.4	191	20	AAx56527	Human ORFX polynuc	c 368	17	0.4	315	22	AAK21634	Probe #4215 used t
c 296	17	0.4	204	22	AAK45600	Human 1508552 DNA	c 369	17	0.4	315	22	AAK21634	Human genome-deriv
c 297	17	0.4	204	22	AAI51536	Human bone marrow	c 370	17	0.4	315	22	AAK21634	Expressed Sequence
c 298	17	0.4	204	24	ABN19872	Probe #20222 used	c 371	17	0.4	315	22	AAK21634	Human brain expres
c 299	17	0.4	219	24	ABN17498	Human genome-deriv	c 372	17	0.4	315	22	AAK21634	Human T84016 DNA f
c 300	17	0.4	220	20	AAx37720	Human ORFX polynuc	c 373	17	0.4	315	22	AAK21634	Eucalyptus grandis
c 301	17	0.4	240	24	ABQ69578	Human clone 150856	c 374	17	0.4	315	22	AAK21634	

c 375	17	0.4	400	22	ABAl11592	Human nervous syst	c 448	17	0.4	621	23	AA584846	DNA encoding novel
c 376	17	0.4	403	24	ABN94460	Gene #958 used to	c 449	17	0.4	622	23	AA584846	DNA encoding novel
c 377	17	0.4	407	23	ABV34082	Human prostate exp	c 450	17	0.4	628	23	ABV51752	Human prostate exp
c 378	17	0.4	407	23	ABV42947	Human prostate exp	c 451	17	0.4	630	23	ABV32786	Human prostate exp
c 379	17	0.4	413	21	ABV67720	Eucaalyptus grandis	c 452	17	0.4	630	23	ABV41711	Human prostate exp
c 380	17	0.4	414	21	AA675161	Human ORFX ORF716	c 453	17	0.4	634	22	AAH36572	Human colon cancer
c 381	17	0.4	414	24	ABN24044	Human ORFX polynuc	c 454	17	0.4	635	22	AAK61487	Human colon cancer
c 382	17	0.4	419	23	ABV43365	Human prostate exp	c 455	17	0.4	660	22	AAH34031	Human colon cancer
c 383	17	0.4	422	22	AAI89729	Human polynucleoti	c 456	17	0.4	664	24	ABK63253	Rat sequence diffe
c 384	17	0.4	430	20	AA56528	Human R02633 DNA f	c 457	17	0.4	680	24	ABL66536	Lung cancer relate
c 385	17	0.4	435	20	AAV56524	Human T87045 DNA f	c 458	17	0.4	681	22	AAH62737	Shrimp white spot
c 386	17	0.4	455	20	AAH89962	EST clone CS776.	c 459	17	0.4	691	21	AAAG7318	Human UNK5 cDNA, s
c 387	17	0.4	455	22	AAK64245	Human immune/haema	c 460	17	0.4	698	24	ABQ56167	Human ovarian anti
c 388	17	0.4	458	22	AAK32376	Human bone marrow	c 461	17	0.4	702	23	AA564926	DNA encoding novel
c 389	17	0.4	458	22	AAI38221	Probe #6907 used t	c 462	17	0.4	708	23	ABV39163	Human prostate exp
c 390	17	0.4	458	24	AB507159	Human genome-deriv	c 463	17	0.4	711	24	ABN70861	Streptococcus poly
c 391	17	0.4	459	21	AAAC00760	Human secreted pro	c 464	17	0.4	726	20	AAK81767	Nucleotide sequenc
c 392	17	0.4	465	22	AAK34651	Human DNA for a no	c 465	17	0.4	732	22	ABA06883	Human pancreatic c
c 393	17	0.4	465	22	AAK34652	Human DNA for a no	c 466	17	0.4	732	22	AAK87862	Human digestive sy
c 394	17	0.4	468	24	ABL95560	Human angiogenesis	c 467	17	0.4	735	24	ABQ65504	Arabidopsis thalia
c 395	17	0.4	468	24	ABL88267	Human DNA64908 pro	c 468	17	0.4	738	22	ABA09427	Human secreted pro
c 396	17	0.4	470	22	ABA26634	Probe #5100 for ge	c 469	17	0.4	750	22	ABAL1862	Human nervous syst
c 397	17	0.4	470	22	AAI15249	Probe #5182 for ge	c 470	17	0.4	750	22	AA544716	Human full-length
c 398	17	0.4	470	24	ABN61018	Human cancer relat	c 471	17	0.4	751	22	AAK61601	Human immune/haema
c 399	17	0.4	471	22	AAK86326	Human immune/haema	c 472	17	0.4	763	22	ABA46978	Human breast cell
c 400	17	0.4	472	22	AAI81118	Human polynucleoti	c 473	17	0.4	763	22	ABA64859	Human foetal liver
c 401	17	0.4	474	22	ABA57991	Human foetal liver	c 474	17	0.4	763	22	ABA31974	Probe #10440 for g
c 402	17	0.4	474	22	AAK06064	Human brain expres	c 475	17	0.4	763	22	AAK13287	Human brain expres
c 403	17	0.4	474	22	AAK31713	Human bone marrow	c 476	17	0.4	763	22	AAK13022	Human bone marrow
c 404	17	0.4	474	22	AAI37589	Probe #6275 used t	c 477	17	0.4	763	22	AAI19832	Probe #9765 for ge
c 405	17	0.4	474	23	ABV62187	Human prostate exp	c 478	17	0.4	763	22	AAI45027	Probe #13713 used
c 406	17	0.4	474	23	AA565281	DNA encoding novel	c 479	17	0.4	763	22	AAI05548	Probe #5539 used t
c 407	17	0.4	474	24	AB506469	Human genome-deriv	c 480	17	0.4	763	22	ABS13106	Human genome-deriv
c 408	17	0.4	477	24	AAAL24854	Human breast cance	c 481	17	0.4	768	22	AAI94760	Human neuroblastom
c 409	17	0.4	482	22	AAI81248	Human polynucleoti	c 482	17	0.4	777	22	AA522474	Human cDNA encodin
c 410	17	0.4	485	22	AAK32725	Human bone marrow	c 483	17	0.4	784	20	AAV84501	Human secreted pro
c 411	17	0.4	493	23	ABV52718	Human prostate exp	c 484	17	0.4	784	22	ABA83284	Human secreted pro
c 412	17	0.4	495	22	ABAA4865	Human breast cell	c 485	17	0.4	790	20	AAK40026	Prostate cancer as
c 413	17	0.4	495	22	ABA55327	Human foetal liver	c 486	17	0.4	790	22	AAK93330	Human cDNA clone r
c 414	17	0.4	495	22	ABA25060	Probe #3526 for ge	c 487	17	0.4	795	24	ABQ54270	Human ovarian anti
c 415	17	0.4	495	22	AAK03569	Human brain expres	c 488	17	0.4	809	23	ABV11641	Human prostate exp
c 416	17	0.4	495	22	AAI34978	Probe #3664 used t	c 489	17	0.4	838	13	AAQ23896	Encodes human prot
c 417	17	0.4	500	24	ABN15730	Human genome-deriv	c 490	17	0.4	846	21	AAZ37135	Mouse TANGO 125 (T
c 418	17	0.4	504	24	ABQ32852	Human genome-deriv	c 491	17	0.4	846	24	AAAD33537	Mouse T125 (TANGO-
c 419	17	0.4	504	24	ABQ32853	Oligonucleotide fo	c 492	17	0.4	867	21	AAAO5453	Streptococcus pneu
c 420	17	0.4	504	24	ABQ32853	Oligonucleotide fo	c 493	17	0.4	876	24	ABQ54672	Human ovarian anti
c 421	17	0.4	507	21	AA280742	Human colon cancer	c 494	17	0.4	883	23	AA573212	DNA encoding novel
c 422	17	0.4	511	24	ABK53402	Human eosinophil'm	c 495	17	0.4	888	24	ABN68808	Streptococcus poly
c 423	17	0.4	518	23	ABU06511	Drosophila melanog	c 496	17	0.4	889	22	AAAD15851	Soybean submergenc
c 424	17	0.4	523	23	ABV02472	Human prostate exp	c 497	17	0.4	891	23	AA580597	DNA encoding novel
c 425	17	0.4	529	23	AA564716	DNA encoding novel	c 498	17	0.4	891	24	ABQ68119	Listeria monocytog
c 426	17	0.4	529	23	AA584382	DNA encoding novel	c 499	17	0.4	891	24	ABQ70003	Listeria monocytog
c 427	17	0.4	538	22	AA531218	Human cDNA encodin	c 500	17	0.4	905	21	AAAC49459	Arabidopsis thalia
c 428	17	0.4	538	24	ABQ66542	Human polynucleoti	c 501	17	0.4	907	21	AAAC35431	Arabidopsis thalia
c 429	17	0.4	549	22	AAH53510	S. epidermidis ope	c 502	17	0.4	916	22	AAK65354	Human immune/haema
c 430	17	0.4	550	21	AAH74848	Human ORFX ORF403	c 503	17	0.4	924	19	AAV28136	Human junctional a
c 431	17	0.4	552	22	AAI44240	Probe #12926 used	c 504	17	0.4	932	20	AAK30154	Human secreted pro
c 432	17	0.4	552	24	AB512341	Human genome-deriv	c 505	17	0.4	950	21	AA56025	Eucaalyptus grandis
c 433	17	0.4	554	22	ABA60661	Human foetal liver	c 506	17	0.4	974	21	AAAF21745	Human breast and o
c 434	17	0.4	554	22	AAK08943	Human brain expres	c 507	17	0.4	976	22	AA528267	Genomic sequence #
c 435	17	0.4	554	22	AAK34834	Human bone marrow	c 508	17	0.4	984	21	AAAC75800	Human ORFX ORF1355
c 436	17	0.4	563	22	AAI35587	Human musculoskele	c 509	17	0.4	984	22	AAAD13235	Human BCL-X-like p
c 437	17	0.4	567	24	ABQ23170	Oligonucleotide fo	c 510	17	0.4	1006	20	AAK79644	Human LKB1 gene fr
c 438	17	0.4	567	24	ABQ23171	Oligonucleotide fo	c 511	17	0.4	1006	24	ABA03935	Human STK11 gene f
c 439	17	0.4	573	22	AAH70179	Human cervical can	c 512	17	0.4	1006	24	ABA04006	Human STK11 gene f
c 440	17	0.4	580	23	ABV52925	Human prostate exp	c 513	17	0.4	1009	20	AAK21838	F11 antigen coding
c 441	17	0.4	595	22	ABA31451	Probe #9917 for ge	c 514	17	0.4	1045	22	AA526177	Human cDNA encodin
c 442	17	0.4	595	22	AAK12780	Human brain expres	c 515	17	0.4	1069	22	AAAF3125	Human secreted pro
c 443	17	0.4	595	22	AAK38499	Human bone marrow	c 516	17	0.4	1080	22	AA533183	DNA encoding human
c 444	17	0.4	595	22	AAI19304	Probe #9237 for ge	c 517	17	0.4	1090	22	AA526430	Human cDNA encodin
c 445	17	0.4	595	24	AB512572	Human genome-deriv	c 518	17	0.4	1092	22	AA526776	Human genomic DNA
c 446	17	0.4	612	20	AAH88991	EST clone IJ638.	c 519	17	0.4	1092	22	AA526777	Human genomic DNA
c 447	17	0.4	614	22	AAH08838	Human cDNA clone (c 520	17	0.4	1112	22	AA545038	cDNA encoding nove

667	17	0.4	2132	22	AAD13237	Human BCL-X-like p	c 740	17	0.4	2784	23	AAS72946	DNA encoding novel
668	17	0.4	2133	22	AAS26108	Human cDNA encoding	741	17	0.4	2784	23	AAS77185	DNA encoding novel
c 669	17	0.4	2136	22	AAS26530	Human cDNA encoding	c 742	17	0.4	2797	22	AAC84603	Human zF3 protein
c 670	17	0.4	2139	17	AA741704	Lymphocyte specifi	743	17	0.4	2808	16	AA705764	Human E-cadherin c
671	17	0.4	2141	22	AK94509	Human full-length	744	17	0.4	2808	24	ABN89392	Human E-cadherin e
672	17	0.4	2161	23	ABL20939	Drosophila melanog	745	17	0.4	2815	15	AAQ65487	Sequence of human
c 673	17	0.4	2162	22	AD16756	Human novel protei	746	17	0.4	2849	20	ABK36030	Human PRO353 nucle
674	17	0.4	2165	21	AA295313	Human mitochondria	c 747	17	0.4	2849	20	AA800058	Human PRO353 nucle
675	17	0.4	2165	24	ABK84712	Human cDNA differe	c 748	17	0.4	2849	21	AA449566	Human PRO353 cDNA
676	17	0.4	2165	24	ABN97328	Gene #3826 used to	c 749	17	0.4	2849	22	AA821429	Human cDNA sequenc
677	17	0.4	2179	23	ABL09929	Drosophila melanog	c 750	17	0.4	2849	22	AA870113	Nucleotide sequenc
c 678	17	0.4	2181	17	AA112500	Non-infective tick	c 751	17	0.4	2859	23	ABL04595	Drosophila melanog
c 679	17	0.4	2181	19	AAV59451	Plasmid construct	c 752	17	0.4	2868	21	AA958525	Human metalloprote
680	17	0.4	2206	24	ABK24325	DNA encoding human	753	17	0.4	2898	24	AB199437	Mouse ischaemic co
c 681	17	0.4	2214	22	AAH68404	C glutamicum codin	c 754	17	0.4	2948	23	ABL15091	Drosophila melanog
c 682	17	0.4	2249	23	ABL29233	Human prostate exp	755	17	0.4	2948	24	AA145975	Human macro protei
c 683	17	0.4	2253	23	ABL29233	Drosophila melanog	c 756	17	0.4	2952	22	AD16759	Human novel protei
684	17	0.4	2259	22	AAS22473	Human cDNA encoding	c 757	17	0.4	2976	23	AA580195	DNA encoding novel
685	17	0.4	2260	20	AA241314	Human normal ovari	758	17	0.4	2976	22	AA158266	Human polynucleoti
c 686	17	0.4	2270	21	AA75892	Human ORFX ORF1447	c 759	17	0.4	3035	23	ABL05379	Drosophila melanog
c 687	17	0.4	2276	20	AA355550	DNA encoding a F-b	760	17	0.4	3105	22	AAH52200	Human AFP protein
c 688	17	0.4	2276	22	AA355550	Human full length	761	17	0.4	3108	24	ABA04428	Human intracellular
689	17	0.4	2277	22	AA844459	Interleukin-1 rece	762	17	0.4	3143	24	AD37608	Human cDNA sequenc
690	17	0.4	2288	21	AA561602	DNA encoding an in	c 763	17	0.4	3185	23	AA591627	DNA encoding novel
691	17	0.4	2293	22	AA157800	Human polynucleoti	c 764	17	0.4	3196	22	AAH98392	Human EST-derived
692	17	0.4	2295	22	AAS22788	Human cDNA encoding	765	17	0.4	3198	23	ABL15738	Drosophila melanog
c 693	17	0.4	2332	22	AAH17008	Human cDNA sequenc	766	17	0.4	3199	23	ABL21330	Drosophila melanog
c 694	17	0.4	2337	22	AA71634	Corynebacterium gl	767	17	0.4	3221	23	ABL06510	Drosophila melanog
c 695	17	0.4	2343	22	AA94524	Human full-length	768	17	0.4	3232	22	AAH54722	S. epidermidis gen
696	17	0.4	2362	20	AA200357	Nucleotide sequenc	c 769	17	0.4	3276	19	AAV59791	Human secreted pro
697	17	0.4	2362	22	AA89019	Human FATP2 coding	770	17	0.4	3376	20	AAK84141	Intron sequence of
c 698	17	0.4	2368	17	AA742128	Neisseria gonorrho	c 771	17	0.4	3386	24	ABK84112	Human cDNA differe
c 699	17	0.4	2381	22	AA010202	N. gonorrheae or N	-c 772	17	0.4	3386	24	ABK64551	Human benign prost
c 700	17	0.4	2418	9	AA81162	Encodes Western su	773	17	0.4	3389	23	ABV22266	Human prostate exp
701	17	0.4	2430	22	ABA15899	Human nervous syst	774	17	0.4	3389	23	ABV28103	Human prostate exp
702	17	0.4	2430	22	ABA16361	Human nervous syst	c 775	17	0.4	3413	22	AAK66240	Human immune/haema
703	17	0.4	2443	21	AA288939	Human soluble guan	c 776	17	0.4	3460	19	AAV36957	Nucleotide sequenc
704	17	0.4	2456	23	AA574097	DNA encoding novel	c 777	17	0.4	3460	21	AA234989	Human cell surface
c 705	17	0.4	2461	22	AA575623	Human cell cycle a	c 778	17	0.4	3460	24	ABK84482	Human cDNA differe
706	17	0.4	2471	22	AAK94363	Human full-length	779	17	0.4	3478	23	ABL15826	Drosophila melanog
707	17	0.4	2490	23	AA586902	DNA encoding novel	c 780	17	0.4	3509	22	AAH15777	Human cDNA sequenc
708	17	0.4	2509	23	ABL20287	Drosophila melanog	c 781	17	0.4	3523	23	ABL09879	Drosophila melanog
709	17	0.4	2512	23	AA567345	DNA encoding novel	c 782	17	0.4	3549	23	ABL10635	Drosophila melanog
c 710	17	0.4	2529	21	AAAC42940	Arabidopsis thalia	c 783	17	0.4	3597	23	ABL09813	Drosophila melanog
c 711	17	0.4	2532	24	ABK11656	DNA encoding human	c 784	17	0.4	3657	23	ABV25797	Human prostate exp
c 712	17	0.4	2553	24	ABK12900	Human protease PRT	785	17	0.4	3670	22	AA160052	Human polynucleoti
c 713	17	0.4	2553	24	ABK31792	DNA encoding novel	c 786	17	0.4	3703	22	AA876536	Human RecQ5 beta D
c 714	17	0.4	2590	21	AA336239	DNA encoding a bo	c 787	17	0.4	3711	23	ABL26102	Drosophila melanog
c 715	17	0.4	2616	14	AAQ37966	Sequence encoding	c 788	17	0.4	3776	24	ABQ54759	Human ovarian anti
c 716	17	0.4	2617	22	AAS26104	Human cDNA encoding	c 789	17	0.4	3807	23	ABV21514	Human prostate exp
c 717	17	0.4	2627	24	ABL64775	Lung cancer relate	c 790	17	0.4	3807	23	ABV22724	Human prostate exp
c 718	17	0.4	2627	24	ABL65409	Lung cancer relate	c 791	17	0.4	3807	23	ABV24950	Human prostate exp
c 719	17	0.4	2627	24	ABL66513	Lung cancer relate	c 792	17	0.4	3807	23	ABV24965	Human prostate exp
c 720	17	0.4	2627	24	ABL68130	Ovary cancer relat	c 793	17	0.4	3807	23	ABV25855	Human prostate exp
c 721	17	0.4	2627	24	ABL69119	Kidney cancer rela	c 794	17	0.4	3807	23	ABV27332	Human prostate exp
c 722	17	0.4	2631	24	AA596334	Arabidopsis cDNA e	c 795	17	0.4	3807	23	ABV28549	Human prostate exp
c 723	17	0.4	2638	11	AAQ03741	Human myb related	c 796	17	0.4	3900	23	AA564300	DNA encoding novel
c 724	17	0.4	2638	19	AAV20468	Human B-myb oncoge	c 797	17	0.4	4043	23	ABL22462	Drosophila melanog
725	17	0.4	2640	23	ABL15830	Drosophila melanog	c 798	17	0.4	4076	23	ABL02152	Drosophila melanog
726	17	0.4	2641	22	AA161175	Human polynucleoti	c 799	17	0.4	4082	22	AAK79954	Human immune/haema
727	17	0.4	2643	20	AA331982	Sulfated fucose-co	c 800	17	0.4	4082	22	AAK82476	Human immune/haema
728	17	0.4	2644	24	ABN59851	Novel human codin	c 801	17	0.4	4085	22	AAK79956	Human immune/haema
729	17	0.4	2650	23	ABL29347	Drosophila melanog	c 802	17	0.4	4085	22	AAK82478	Human immune/haema
c 730	17	0.4	2651	19	AAV52349	Streptococcus pneu	c 803	17	0.4	4086	22	AAH49373	Human metalloprote
731	17	0.4	2654	23	ABL19179	Drosophila melanog	c 804	17	0.4	4086	22	AAK79955	Human immune/haema
c 732	17	0.4	2656	24	AA594947	Human DNA sequenc	c 805	17	0.4	4086	22	AAK82477	Human immune/haema
c 733	17	0.4	2661	22	AA159389	Human polynucleoti	c 806	17	0.4	4123	24	ABK50067	DNA encoding mouse
c 734	17	0.4	2735	23	AA585641	DNA encoding novel	c 807	17	0.4	4148	21	AAA46556	cDNA sequence enco
c 735	17	0.4	2748	23	ABL09795	Drosophila melanog	c 808	17	0.4	4148	24	ABK72679	DNA encoding enco
736	17	0.4	2749	23	ABL34423	Human immune syste	c 809	17	0.4	4260	23	AA574826	DNA encoding novel
737	17	0.4	2754	23	AA568677	DNA encoding novel	c 810	17	0.4	4270	24	ABN95930	Gene #2428 used to
c 738	17	0.4	2769	20	AA224813	Human adipocyte-sp	811	17	0.4	4270	24	ABL64082	Breast cancer rela
c 739	17	0.4	2769	21	AAA64114	DNA encoding a hum	812	17	0.4	4283	22	AAS25994	Human cDNA encodin

c 813	17	0.4	4321	23	ABL17250	Drosophila melanog	886	17	0.4	7131	24	AAS61361	Human gene regulat
c 814	17	0.4	4358	24	AAI67895	Nucleotide sequenc	887	17	0.4	7131	24	ABK31451	Signal transductio
c 815	17	0.4	4361	23	ABL04510	Drosophila melanog	888	17	0.4	7306	24	ABL33636	Human immune syste
c 816	17	0.4	4406	22	ABL07188	Human reproductive	c 889	17	0.4	7308	23	ABL10634	Drosophila melanog
c 817	17	0.4	4411	24	AAS94987	Human DNA sequence	890	17	0.4	7374	22	ABA08211	Human ovarian and
c 818	17	0.4	4529	20	AAK80829	Human secreted pro	891	17	0.4	7374	22	AAI07519	RNA encoding enter
c 819	17	0.4	4529	21	AAZ52477	Human secreted pro	892	17	0.4	7399	12	AAQ11816	Human immune/haema
c 820	17	0.4	4569	22	AAI60358	Human polynucleoti	c 893	17	0.4	7908	22	AAK66239	Human immune/haema
c 821	17	0.4	4569	22	AAI60359	Human polynucleoti	894	17	0.4	7984	24	ABK91610	Modified HIV prote
c 822	17	0.4	4569	22	AAI60360	Human polynucleoti	c 895	17	0.4	8004	23	ABL29346	Drosophila melanog
c 823	17	0.4	4574	22	ABA17564	Human polynucleoti	c 896	17	0.4	8033	23	ABL19178	Drosophila melanog
c 824	17	0.4	4587	23	ABV22996	Human nervous syst	897	17	0.4	8067	23	ABL15758	Drosophila melanog
c 825	17	0.4	4587	23	ABV22996	Human prostate exp	898	17	0.4	8150	23	ABL20938	Drosophila melanog
c 826	17	0.4	4679	22	AAK45282	Human encoding nove	899	17	0.4	8265	22	AAK42165	Genomic sequence #
c 827	17	0.4	4736	22	AAK45094	cDNA encoding nove	c 900	17	0.4	8340	23	ABL15624	Drosophila melanog
c 828	17	0.4	4900	22	AAK29462	Human Slit-1 cDNA.	901	17	0.4	8399	23	ABL20286	Drosophila melanog
c 829	17	0.4	4950	20	AAK89162	Human Slit-2 prote	902	17	0.4	8401	23	ABL09928	Drosophila melanog
c 830	17	0.4	4950	20	AAK14979	Nucleic acid encod	c 903	17	0.4	8486	24	ABL34444	Human immune syste
c 831	17	0.4	5045	23	ABL09794	Drosophila melanog	904	17	0.4	8789	22	AAK82834	Human immune/haema
c 832	17	0.4	5050	22	ABA16961	Human nervous syst	905	17	0.4	8922	22	AAK87069	Human immune/haema
c 833	17	0.4	5053	24	ABK88916	Human aggreganase	906	17	0.4	9280	24	ABK91615	Modified HIV prote
c 834	17	0.4	5055	23	ABL29232	Drosophila melanog	907	17	0.4	9285	24	ABK91608	Modified HIV prote
c 835	17	0.4	5086	22	AAK07355	Human KIAA0211 cDN	908	17	0.4	9524	24	ABK39993	Human chemically p
c 836	17	0.4	5086	24	ABK83718	Human cDNA differe	909	17	0.4	9524	24	ABL32838	Human immune syste
c 837	17	0.4	5102	23	ABL06280	Drosophila melanog	c 910	17	0.4	9543	22	AAK36312	Human musculoskele
c 838	17	0.4	5107	24	ABL33066	Human immune syste	c 911	17	0.4	9703	22	AAS33480	DNA encoding human
c 839	17	0.4	5125	22	ABA21120	Human nervous syst	c 912	17	0.4	9881	24	ABL54354	Chemically treated
c 840	17	0.4	5125	22	AAK67700	Human immune/haema	c 913	17	0.4	9904	23	ABL04594	Drosophila melanog
c 841	17	0.4	5125	22	AAK78883	Human immune/haema	914	17	0.4	9910	22	ABA18525	Human nervous syst
c 842	17	0.4	5206	23	ABL05378	Drosophila melanog	915	17	0.4	10401	22	AAS42164	Genomic sequence #
c 843	17	0.4	5228	20	AAV81384	Nucleic acid encod	916	17	0.4	10461	20	AAK20553	Polynucleotide seq
c 844	17	0.4	5233	21	AAK63530	Slit-2 protein cod	c 917	17	0.4	11141	17	AAK44469	Tick-borne enceph
c 845	17	0.4	5248	23	ABL05198	Drosophila melanog	c 918	17	0.4	11141	17	AAK15820	TBE virus strain N
c 846	17	0.4	5250	22	AAK52371	Human polynucleoti	919	17	0.4	11743	23	ABL15248	Drosophila melanog
c 847	17	0.4	5325	23	ABL09672	Drosophila melanog	c 920	17	0.4	12177	24	ABL32650	Human immune syste
c 848	17	0.4	5375	22	AAK69007	Human immune/haema	c 921	17	0.4	12263	24	AAK84514	Human cDNA differe
c 849	17	0.4	5605	24	ABK88915	Human aggreganase	c 922	17	0.4	12351	22	AAS41943	Genomic sequence #
c 850	17	0.4	5609	23	ABL15090	Drosophila melanog	c 923	17	0.4	12410	22	AAS41942	Genomic sequence #
c 851	17	0.4	5646	23	ABL02910	Drosophila melanog	c 924	17	0.4	12412	22	AAS41941	Genomic sequence #
c 852	17	0.4	5724	24	ABK10762	DNA encoding human	c 925	17	0.4	12870	23	ABL14952	Drosophila melanog
c 853	17	0.4	5785	22	AAK42123	Genomic sequence #	c 926	17	0.4	12881	22	AAK85861	Human immune/haema
c 854	17	0.4	5785	22	AAK42125	Genomic sequence #	c 927	17	0.4	13332	23	ABL09812	Drosophila melanog
c 855	17	0.4	5804	22	AAK86620	Human immune/haema	c 928	17	0.4	13469	22	AAS46593	Tumour suppressor
c 856	17	0.4	5804	22	AAK63443	Human ADAMTS-9 cDN	c 929	17	0.4	13587	23	ABL11308	Drosophila melanog
c 857	17	0.4	5808	24	AAK597179	Human metalloprote	c 930	17	0.4	14372	23	ABL09878	Drosophila melanog
c 858	17	0.4	5819	22	AAK42127	Genomic sequence #	c 931	17	0.4	14448	22	AAK71815	Human immune/haema
c 859	17	0.4	5844	24	ABL49389	Human polynucleoti	c 932	17	0.4	14448	22	AAK73111	Human immune/haema
c 860	17	0.4	5849	24	ABK17335	Human Rb- interacti	c 933	17	0.4	14448	22	AAK87580	Human immune/haema
c 861	17	0.4	5868	21	AAA60104	Human RIZ allele D	c 934	17	0.4	14448	22	AAI62922	Human genomic DNA
c 862	17	0.4	5868	21	AAA60123	Human RIZ allele E	c 935	17	0.4	15081	23	ABL30084	Drosophila melanog
c 863	17	0.4	5868	21	AAA60124	Human mutant RIZ a	c 936	17	0.4	15270	22	AAK74017	Human immune/haema
c 864	17	0.4	5868	24	AAK18781	DNA sequence encod	c 937	17	0.4	15270	22	AAK85254	Human immune/haema
c 865	17	0.4	5874	23	AAS85420	DNA encoding nove	c 938	17	0.4	15297	24	ABA94501	Human proto-oncog
c 866	17	0.4	5949	24	AAK98771	Human ADAMTS-SI nu	c 939	17	0.4	15453	22	AAK29339	Atopy related gene
c 867	17	0.4	5986	22	AAS45439	Chemically pretrea	c 940	17	0.4	15809	22	AAK68321	Human immune/haema
c 868	17	0.4	5986	24	ABK28288	DNA transcription	c 941	17	0.4	16310	21	AAK21086	Human low adenosin
c 869	17	0.4	5990	22	AAK63449	Human ADAMTS-9 alt	c 942	17	0.4	16310	21	AAA34964	Human adenosine re
c 870	17	0.4	6013	24	ADK30577	Human ADAMTS-9 alt	c 943	17	0.4	16348	22	AAK76522	Human immune/haema
c 871	17	0.4	6060	24	ABK10763	Human ADAMTS-9 alt	c 944	17	0.4	16348	22	AAK76522	Human immune/haema
c 872	17	0.4	6084	24	ABL33699	Human ADAMTS-9 alt	c 945	17	0.4	16605	22	AAK85489	Human immune/haema
c 873	17	0.4	6126	24	ABL33445	Human immune syste	c 946	17	0.4	16605	22	AAK87049	Human immune/haema
c 874	17	0.4	6126	24	ABK31335	Signal transductio	c 947	17	0.4	16989	23	ABL19384	Drosophila melanog
c 875	17	0.4	6216	22	AAK69008	Human immune/haema	c 948	17	0.4	17462	22	ABA15655	Human nervous syst
c 876	17	0.4	6223	22	AAI58572	Human polynucleoti	c 949	17	0.4	17634	21	AAK21087	Human low adenosin
c 877	17	0.4	6304	22	AAI58574	Human polynucleoti	c 950	17	0.4	17634	21	AAA34965	Human adenosine re
c 878	17	0.4	6373	22	ABA14642	Human nervous syst	c 951	17	0.4	18595	22	AAK33411	DNA encoding human
c 879	17	0.4	6381	22	ABA14641	Human nervous syst	c 952	17	0.4	19125	22	AAK36440	Human musculoskele
c 880	17	0.4	6382	22	AAI58573	Human polynucleoti	c 953	17	0.4	19125	19	AAV52232	Streptococcus pneu
c 881	17	0.4	6535	22	AAS46470	Tumour suppressor	c 954	17	0.4	19969	22	AAK65309	Human immune/haema
c 882	17	0.4	6596	19	AAV57377	Maize female-prefe	c 955	17	0.4	20598	22	AAK24851	Nucleotide sequenc
c 883	17	0.4	6834	22	AAI05314	Human reproductive	c 956	17	0.4	21423	22	AAK36470	Human musculoskele
c 884	17	0.4	6834	22	AAS28901	Human immunoglobul	c 957	17	0.4	21477	22	AAK66626	Human immune/haema
c 885	17	0.4	7131	24	ABL70428	Chemically treated	c 958	17	0.4	21480	22	AAK66625	Human immune/haema

Db 651 TGTTCGAAACCCAGTCGCAAAAAGATATTTTAAATAGAGGCTGCCGACTTCATTGAAGAG 710
QY 481 CTATCCTCCCTTTTCAATCCACAGCTCGAAAAGATTAGACCTCGTGCCTGCAAAAAC 540
Db 711 CTATCCTCCCTTTTCAATCCACAGCTCGAAAAGATTAGACCTCGTGCCTGCAAAAAC 770
QY 541 CACAAGAGTAACTGGAATCTCAAAACAAAGTTATGCGAGAAACAGCTCCAGTTTCTCA 600
Db 771 CACAAGAGTAACTGGAATCTCAAAACAAAGTTATGCGAGAAACAGCTCCAGTTTCTCA 830
QY 601 GATCTGTCAAGAGACGAGAAAGATCTTCTGTTCCTATCCCTATCCCTGCGGATACCAG 660
Db 831 GATCTGTCAAGAGACGAGAAAGATCTTCTGTTCCTATCCCTATCCCTGCGGATACCAG 890
QY 661 GATAATGAAGTGAATCACGCCCTGGAAACAGCAGGAGGCCAAGAGGCTGAAGCGGAGCAG 720
Db 891 GATAATGAAGTGAATCACGCCCTGGAAACAGCAGGAGGCCAAGAGGCTGAAGCGGAGCAG 950
QY 721 GCTGCCAGTGAGCGGCTGGTGGAGACACTACACCAAGGCTTCCCTTTCATCTCTGTAC 780
Db 951 GCTGCCAGTGAGCGGCTGGTGGAGACACTACACCAAGGCTTCCCTTTCATCTCTGTAC 1010
QY 781 TATGAGAACCTCTGGGCAACCTCCCGGTTCACTCAAAAGTTAGCGGACGAGAGTT 840
Db 1011 TATGAGAACCTCTGGGCAACCTCCCGGTTCACTCAAAAGTTAGCGGACGAGAGTT 1070
QY 841 CCAGAAGGAACCTCGAGTTCAGTTGGATTGCATAGTGTAGGAATTCACACCTCAAGTA 900
Db 1071 CCAGAAGGAACCTCGAGTTCAGTTGGATTGCATAGTGTAGGAATTCACACCTCAAGTA 1130
QY 901 AGGTGCTACTGTGAAGCAAGGAGCTTGAAAATTTCCCGAGATATTCACATCGTCCAGGCA 960
Db 1131 AGGTGCTACTGTGAAGCAAGGAGCTTGAAAATTTCCCGAGATATTCACATCGTCCAGGCA 1190
QY 961 GGAAATCTGCATCACTGACCTGCGGAGCCTTTGAGAGGACACAGACGCTATTCC 1020
Db 1191 GGAAATCTGCATCACTGACCTGCGGAGCCTTTGAGAGGACACAGACGCTATTCC 1250
QY 1021 TGTCTTGTCTTAACTATCTATGGACAGATTGACATCTGCTGAGATTTATATAGAAGG 1080
Db 1251 TGTCTTGTCTTAACTATCTATGGACAGATTGACATCTGCTGAGATTTATATAGAAGG 1310
QY 1081 GTTCTTCTTCTGACTCAGAAGCGACCCCTAACAAAGGAAGAGATGAATCGAATCCAGAAG 1140
Db 1311 GTTCTTCTTCTGACTCAGAAGCGACCCCTAACAAAGGAAGAGATGAATCGAATCCAGAAG 1370
QY 1141 CCAATGAGGTGTCATCTCTCCACTACTCTGCGAGTCAATCTCCAGCAGTACCCCAA 1200
Db 1371 CCAATGAGGTGTCATCTCTCCACTACTCTGCGAGTCAATCTCCAGCAGTACCCCAA 1430
QY 1201 GCCCAGCATTTGGTGCCCAACCTCGTGTGGCAACCATCCAGCAGTGTGAGAGCCCCACC 1260
Db 1431 GCCCAGCATTTGGTGCCCAACCTCGTGTGGCAACCATCCAGCAGTGTGAGAGCCCCACC 1490
QY 1261 AATTACTTCAGGGATTGGATGGAACCTATCATTTCCAGCTCCTGTGTTTACAAAGATG 1320
Db 1491 AATTACTTCAGGGATTGGATGGAACCTATCATTTCCAGCTCCTGTGTTTACAAAGATG 1550
QY 1321 CTACAAAATTTGTCAGCTTCTGAGGCTCAGCTGGTGTCTTTGAATCCAGAGTAAAGGA 1380
Db 1551 CTACAAAATTTGTCAGCTTCTGAGGCTCAGCTGGTGTCTTTGAATCCAGAGTAAAGGA 1610
QY 1381 GCTCCATCTCTCAAGTTGAGTGGTATAGAGAAGGAGCTTTAATAGAAGATTTCTCCAGAT 1440
Db 1611 GCTCCATCTCTCAAGTTGAGTGGTATAGAGAAGGAGCTTTAATAGAAGATTTCTCCAGAT 1670
QY 1441 TTTAGGATTTTACAGAAAAACCTTCGATCGATCGGAGGCCAG 1483
Db 1671 TTTAGGATTTTACAGAAAAACCTTCGATCGGAGGCCAG 1713

RESULT 2
AAS74798

ID AAS74798 standard; cDNA; 1228 BP.
XX AC AAS74798;
XX XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10602.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG10611.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1: SEQ ID No 10602; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94584 represent novel human
CC amino acid coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1228 BP; 323 A; 319 C; 314 G; 272 T; 0 other;

Query Match 26.6%; Score 1056; DB 23; Length 1228;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1226; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2737 ATGAATGAATAGAGTTTCCTTGGAAACGCTACTCTTGTGATGAATCAGATGATGAATT 2796
Db 1 ATGAATGAATAGAGTTTCCTTGGAAACGCTACTCTTGTGATGAATCAGATGATGAATT 60
QY 2797 CAACATCATCAGATCCCCACGGCGCAAGTGTATTCCTCCCATCTTGCACAGACTCAAG 2856
Db 61 CAACATCATCAGATCCCCACGGCGCAAGTGTATTCCTCCCATCTTGCACAGACTCAAG 120
QY 2857 CACTTCGGGTCACAGAAGGCTCTCCAGTCACATTCACCTGCAAAATTTGTTGGGATACCT 2916

Db 121 CACTTCGGGTACAGAAGGCTCTCCAGTTACATTACCTCAAAATTTGTTGGATACCT 180
Qy 2917 GTTCCAAAGGTTTACTGGTTCAAGATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGC 2976
Db 181 GTTCCAAAGGTTTACTGGTTCAAGATGGGAAGCAGATTTCTAAGAGAAATGAGCACTGC 240
Qy 2977 AAAATGAGGCGAGAGAGATGGGACATGCTCTCTGCACATTGAATCCACTACCAGTGAT 3036
Db 241 AAAATGAGGCGAGAGAGATGGGACATGCTCTCTGCACATTGAATCCACTACCAGTGAT 300
Qy 3037 GACGATGGCAACTACACCATCATGGCAGCCAAACCCCGAGGGGAGAATCAGCTGTTCTGGC 3096
Db 301 GACGATGGCAACTACACCATCATGGCAGCCAAACCCCGAGGGGAGAATCAGCTGTTCTGGC 360
Qy 3097 CACTTGATGTTACAAAGTTTGGCCATTTCGAGTCGGCTACCTCTGCTGGTCACTCTCAC 3156
Db 361 CACTTGATGTTACAAAGTTTGGCCATTTCGAGTCGGCTACCTCTGCTGGTCACTCTCAC 420
Qy 3157 AGGGGAAGATCCCGAGTGCAGGAAGAGACAAAGAGCCCTTACAGAGACGCTTTTCCGA 3216
Db 421 AGGGGAAGATCCCGAGTGCAGGAAGAGACAAAGAGCCCTTACAGAGACGCTTTTCCGA 480
Qy 3217 CCACATTTCTGTCAGGCTCTCTGGGATATGTTAGCTCATGAGGGGCGCCTCTGTCGGCTG 3276
Db 481 CCACATTTCTGTCAGGCTCTCTGGGATATGTTAGCTCATGAGGGGCGCCTCTGTCGGCTG 540
Qy 3277 GACTGTAAGTGAAGTGTATACCCGCCCGGAGC-TGACATGGCTACTCAATGGCCAAAC 3335
Db 541 GACTGTAAGTGAAGTGTATACCCGCCCGGAGCATGACATGGCTACTCAATGGCCAAAC 600
Qy 3336 TGTGCTACCAAGATGCTTCCCAAGATGCTGGTCAGGAGACCGAGTCCACTCTCTGCT 3395
Db 601 TGTGCTACCAAGATGCTTCCCAAGATGCTGGTCAGGAGACCGAGTCCACTCTCTGCT 660
Qy 3396 CATTGACCCACTACTCAGCGCGACGAGGACCTATAAGTGCATCGCTACCAACAAAAC 3455
Db 661 CATTGACCCACTACTCAGCGCGACGAGGACCTATAAGTGCATCGCTACCAACAAAAC 720
Qy 3456 CGGCGCAATTTCTTTAGTCTGGAGCTCTCTGTAGTAGCCAAAGAGGTGAAGAAGCAC 3515
Db 721 CGGCGCAATTTCTTTAGTCTGGAGCTCTCTGTAGTAGCCAAAGAGGTGAAGAAGCAC 780
Qy 3516 TGTGATCCTGGAGAACTACAGAACTCGGTGTTCGCCAAGGCCACCCGCTGAGACTGGA 3575
Db 781 TGTGATCCTGGAGAACTACAGAACTCGGTGTTCGCCAAGGCCACCCGCTGAGACTGGA 840
Qy 3576 GTGCCCGGTGATAGGCATGCCCCACCTGTGTTCTACTGGAAGAAAGACAATGAGACCAT 3635
Db 841 GTGCCCGGTGATAGGCATGCCCCACCTGTGTTCTACTGGAAGAAAGACAATGAGACCAT 900
Qy 3636 CCCTTGACACAGAGAGGATCAGTATGCACAGGACACACAGGGTATGCTGCCCTTCT 3695
Db 901 CCCTTGACACAGAGAGGATCAGTATGCACAGGACACACAGGGTATGCTGCCCTTCT 960
Qy 3696 CATTACCCAGCCAAAGAAATCAGACGCTGGATGGTACACGTTGTTCAGCCAAAGATGAAGC 3755
Db 961 CATTACCCAGCCAAAGAAATCAGACGCTGGATGGTACACGTTGTTCAGCCAAAGATGAAGC 1020
Qy 3756 CGGCATCGTGTGTCGACTGCCAGGCTGGATATACGCTCAGTGGCACCACATCAGATCCC 3815
Db 1021 CGGCATCGTGTGTCGACTGCCAGGCTGGATATACGCTCAGTGGCACCACATCAGATCCC 1080
Qy 3816 ACCGCCATGCTGTCCGGCCCGAGTGCAGTCGCTACGGATCTCTCAGCAGTAAGGACT 3875
Db 1081 ACCGCCATGCTGTCCGGCCCGAGTGCAGTCGCTACGGATCTCTCAGCAGTAAGGACT 1140
Qy 3876 TGACATATTTTCTGCCCTTTTCTCCATGGAAGACACGATGGTGTATTCATGCTCTCTCG 3935
Db 1141 TGACATATTTTCTGCCCTTTTCTCCATGGAAGACACGATGGTGTATTCATGCTCTCTCG 1200
Qy 3936 GAGTGTAGTGAGAGTGATGAACCTTTAA 3963
|||||

Db 1201 GAGTGTAGTGGAGAGTGATGAACCTTTAA 1228
RESULT 3
ABA06552
ID ABA06552 standard; cDNA; 2059 BP.
XX
AC ABA06552;
XX
XX 10-JAN-2002 (first entry)
XX Human cDNA SEQ ID NO: 218.
DE
DE
XX Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation; ss.
XX
OS Homo sapiens.
XX
XX WO200154474-A2.
PN
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01349.
XX
XX 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225288P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226686P.
PR 23-AUG-2000; 2000US-227182P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.

QY	3657	CAGTATGCACCCAGGACACAAACAGGGTATGCCTGCTTCTCATTCAGCCAGCCAAAGAAATC	3716	PR	14-AUG-2000;	2000US-0225267.
				PR	14-AUG-2000;	2000US-0225268.
Db	489	CAGTATGCACCCAGGACACAAACAGGGTATGCCTGCTTCTCATTCAGCCAGCCAAAGAAATC	548	PR	14-AUG-2000;	2000US-0225270.
				PR	14-AUG-2000;	2000US-0225447.
QY	3717	AGACGCTGGATGGTACACGTTGTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGCACTGC	3776	PR	14-AUG-2000;	2000US-0225757.
				PR	14-AUG-2000;	2000US-0225758.
Db	549	AGACGCTGGATGGTACACGTTGTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGCACTGC	608	PR	14-AUG-2000;	2000US-0225759.
				PR	18-AUG-2000;	2000US-0226279.
QY	3777	CAGGCTGGATATACGCTCAGTGGCCACCATCAGATCCCACCCGCCATGTGTCTCCGGCC	3836	PR	22-AUG-2000;	2000US-0226681.
				PR	22-AUG-2000;	2000US-0226868.
Db	609	CAGGCTGGATATACGCTCAGTGGCCACCATCAGATCCCACCCGCCATGTGTCTCCGGCC	668	PR	22-AUG-2000;	2000US-0227182.
				PR	23-AUG-2000;	2000US-0227009.
QY	3837	CAGTGGCAGTCGCTACCGATCTCTCACAGTAAAGGACTTGGACATATTTTCTGCCTTTTC	3896	PR	30-AUG-2000;	2000US-0228924.
				PR	01-SEP-2000;	2000US-0229287.
Db	669	CAGTGGCAGTCGCTACCGATCTCTCACAGTAAAGGACTTGGACATATTTTCTGCCTTTTC	728	PR	01-SEP-2000;	2000US-0229343.
				PR	01-SEP-2000;	2000US-0229344.
QY	3897	CTCCATGGAAGCACGATGGTATTCATGCTCTTCGCGAGTGTAGTGAGAGTGATGA	3956	PR	01-SEP-2000;	2000US-0229345.
				PR	05-SEP-2000;	2000US-0229509.
Db	729	CTCCATGGAAGCACGATGGTATTCATGCTCTTCGCGAGTGTAGTGAGAGTGATGA	788	PR	05-SEP-2000;	2000US-0229513.
				PR	06-SEP-2000;	2000US-0230437.
QY	3957	ACTTTAA 3963		PR	06-SEP-2000;	2000US-0230438.
				PR	08-SEP-2000;	2000US-0230438.
Db	789	ACTTTAA 795		PR	08-SEP-2000;	2000US-0231242.
				PR	08-SEP-2000;	2000US-0231243.
RESULT 4				PR	08-SEP-2000;	2000US-0231244.
AAS28777				PR	08-SEP-2000;	2000US-0231413.
ID AAS28777 standard; cDNA; 2232 BP.				PR	08-SEP-2000;	2000US-0231414.
XX AC AAS28777;				PR	08-SEP-2000;	2000US-0231414.
XX DT 07-NOV-2001 (first entry)				PR	08-SEP-2000;	2000US-0232081.
XX DE Human immunoglobulin encoding cDNA SEQ ID No 23.				PR	08-SEP-2000;	2000US-0232081.
XX KW Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;				PR	12-SEP-2000;	2000US-0231968.
XX KW antisense therapy; gene therapy; neurological disorder; renal disorder;				PR	14-SEP-2000;	2000US-0232397.
XX KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;				PR	14-SEP-2000;	2000US-0232398.
XX KW reproductive disorder; immune system disorder; proliferative disorder;				PR	14-SEP-2000;	2000US-0232399.
XX KW muscular disorder.				PR	14-SEP-2000;	2000US-0232400.
XX OS Homo sapiens.				PR	14-SEP-2000;	2000US-0232401.
XX PN WO200155315-A2.				PR	14-SEP-2000;	2000US-0232401.
XX PD 02-AUG-2001.				PR	14-SEP-2000;	2000US-0233063.
XX PF 17-JAN-2001; 2001WO-US01326.				PR	14-SEP-2000;	2000US-0233064.
XX PR 31-JAN-2000; 2000US-0179065.				PR	14-SEP-2000;	2000US-0233065.
PR 04-FEB-2000; 2000US-0180828.				PR	21-SEP-2000;	2000US-0233065.
PR 24-FEB-2000; 2000US-0184664.				PR	21-SEP-2000;	2000US-0234223.
PR 02-MAR-2000; 2000US-0186350.				PR	21-SEP-2000;	2000US-0234274.
PR 16-MAR-2000; 2000US-0189874.				PR	25-SEP-2000;	2000US-0234997.
PR 17-MAR-2000; 2000US-0190076.				PR	25-SEP-2000;	2000US-0234997.
PR 18-APR-2000; 2000US-0198123.				PR	25-SEP-2000;	2000US-0234998.
PR 19-MAY-2000; 2000US-0205515.				PR	26-SEP-2000;	2000US-0235484.
PR 07-JUN-2000; 2000US-0209467.				PR	27-SEP-2000;	2000US-0235834.
PR 28-JUN-2000; 2000US-0214886.				PR	27-SEP-2000;	2000US-0235836.
PR 30-JUN-2000; 2000US-0215135.				PR	29-SEP-2000;	2000US-0236327.
PR 07-JUL-2000; 2000US-0216647.				PR	29-SEP-2000;	2000US-0236367.
PR 07-JUL-2000; 2000US-0216880.				PR	29-SEP-2000;	2000US-0236368.
PR 11-JUL-2000; 2000US-0217487.				PR	29-SEP-2000;	2000US-0236369.
PR 11-JUL-2000; 2000US-0217496.				PR	02-OCT-2000;	2000US-0236802.
PR 14-JUL-2000; 2000US-0218290.				PR	02-OCT-2000;	2000US-0237037.
PR 26-JUL-2000; 2000US-0220963.				PR	02-OCT-2000;	2000US-0237038.
PR 26-JUL-2000; 2000US-0220964.				PR	02-OCT-2000;	2000US-0237039.
PR 14-AUG-2000; 2000US-0224518.				PR	02-OCT-2000;	2000US-0237040.
PR 14-AUG-2000; 2000US-0224519.				PR	13-OCT-2000;	2000US-0239935.
PR 14-AUG-2000; 2000US-0225213.				PR	13-OCT-2000;	2000US-0239937.
PR 14-AUG-2000; 2000US-0225214.				PR	20-OCT-2000;	2000US-0240960.
PR 14-AUG-2000; 2000US-0225266.				PR	20-OCT-2000;	2000US-0241221.
				PR	20-OCT-2000;	2000US-0241785.
				PR	20-OCT-2000;	2000US-0241786.
				PR	20-OCT-2000;	2000US-0241787.
				PR	20-OCT-2000;	2000US-0241808.
				PR	20-OCT-2000;	2000US-0241809.
				PR	20-OCT-2000;	2000US-0241826.
				PR	01-NOV-2000;	2000US-0244617.
				PR	08-NOV-2000;	2000US-0246474.
				PR	08-NOV-2000;	2000US-0246475.
				PR	08-NOV-2000;	2000US-0246476.
				PR	08-NOV-2000;	2000US-0246477.
				PR	08-NOV-2000;	2000US-0246478.
				PR	08-NOV-2000;	2000US-0246523.
				PR	08-NOV-2000;	2000US-0246524.
				PR	08-NOV-2000;	2000US-0246525.

08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246509.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
08-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-02559678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-457725/49.
P-PSDB; AAU17989.
Isolated novel immunoglobulin polypeptide for monitoring the presence
and progression of diseases and for diagnosis -
Claim 1; SEQ ID No 23; 551pp; English.
Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the
immunoglobulin polypeptides of the invention, and primers for the
polynucleotides. The polynucleotides and polypeptides can be used to
diagnose a pathological condition or a susceptibility to a pathological
condition in a subject by determining the presence or absence of a
mutation in a DNA sequence or determining the presence or amount of
expression of the protein. Alternatively the identification of a binding
partner to a sequence allows determination of changes in protein
activity. The sequences can be used as research tools for receptors or
other signal transduction pathway proteins that interact with the
polypeptides of the invention and can be used to treat, prevent or
diagnose various types of disorders such as neurological disorders,
cardiovascular disorders, gastrointestinal disorders, reproductive
disorders, immune system disorders, renal disorders, muscular disorders,
pulmonary disorders, proliferative disorders and cancer.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 2232 BP; 617 A; 508 C; 528 G; 569 T; 10 other;

Query Match
Best Local Similarity 17.3%; Score 685; DB 22; Length 2232;
Matches 785; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3177 AGAAAGAGACAAAGAGCCCTACAGGAAGCTTTTCCGACACACATTTCTCTCAGAGCTCC 3236
DB 2 AGAAAGAGACAAAGAGCCCTACAGGAAGCTTTTCCGACACACATTTCTCTCAGAGCTCC 61
QY 3237 TGGGGATATGGTAGCTCATGAGGGGGCCCTCTGTGGCTGGACTGTAAGGTGAGTGGTTT 3296
DB 62 TGGGGATATGGTAGCTCATGAGGGGGCCCTCTGTGGCTGGACTGTAAGGTGAGTGGTTT 121
QY 3297 ACCGGCCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTCTACAGATGCCCTCCCA 3356
DB 122 ACCGGCCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTCTACAGATGCCCTCCCA 181
QY 3357 CAAGATGCTGGTCAGGGAGACCGGAGTCCACTCTCTGTCTCAATGACCCACTCCTCAGCG 3416
DB 182 CAAGATGCTGGTCAGGGAGACCGGAGTCCACTCTCTGTCTCAATGACCCACTCCTCAGCG 241
QY 3417 CGACGCGAGGACCTATAAGTGCATCGCTACCAACAAACCGGCGAGAAATCTTTTAGTCT 3476
DB 242 CGACGCGAGGACCTATAAGTGCATCGCTACCAACAAACCGGCGAGAAATCTTTTAGTCT 301
QY 3477 GGAGCTCTCTGTAGTCCAAAGAGGTGAAGAAAGCACCTGTGTATCCTGGAGAAACTACA 3536
DB 302 GGAGCTCTCTGTAGTCCAAAGAGGTGAAGAAAGCACCTGTGTATCCTGGAGAAACTACA 361
QY 3537 GAACTGCGGTGTTCGCCAAGGCCACCCCGTGAGACTGAGTGCCTGATAGGCATGCC 3596
DB 362 GAACTGCGGTGTTCGCCAAGGCCACCCCGTGAGACTGAGTGCCTGATAGGCATGCC 421
QY 3597 CCCACCTGTCTTCTACTGGAAGAACACATGAGACCATCCCTTGACACGAGAGAGGAT 3656
DB 422 CCCACCTGTCTTCTACTGGAAGAACACATGAGACCATCCCTTGACACGAGAGAGGAT 481
QY 3657 CAGTATGCACACGAGCACACAGGGTATGCTGCCCTTCTCATTACGACCCCAAGAAATC 3716
DB 482 CAGTATGCACACGAGCACACAGGGTATGCTGCCCTTCTCATTACGACCCCAAGAAATC 541
QY 3717 AGACGCTGGATGGTACACAGCTGTGTACGCCAAGAAATGAAGCGGCGATCGTGTGCACTGC 3776
DB 542 AGACGCTGGATGGTACACAGCTGTGTACGCCAAGAAATGAAGCGGCGATCGTGTGCACTGC 601
QY 3777 CAGGCTGGATATATACGCTCAGTGGCACCACATCAGATCCACCCCATGCTCTCTCGGCC 3836
DB 602 CAGGCTGGATATATACGCTCAGTGGCACCACATCAGATCCACCCCATGCTCTCTCGGCC 661
QY 3837 CAGTGGCAGTTCGCTACGGATCTCTCACCAGTAAAGACTTGACATATTTTCTGCCCTTTC 3896
DB 662 CAGTGGCAGTTCGCTACGGATCTCTCACCAGTAAAGACTTGACATATTTTCTGCCCTTTC 721
QY 3897 CTCCATGGAAGACACGATGGTGTATTCTCTCGAGTGTAGTGGAGAGTGTATGA 3956
DB 722 CTCCATGGAAGACACGATGGTGTATTCTCTCGAGTGTAGTGGAGAGTGTATGA 781
QY 3957 ACTTTAA 3963
DB 782 ACTTTAA 788
RESULT 5
AAH04212
ID AAH04212 standard; cDNA; 855 BP.
XX
AC AAH04212;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:1047.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

```
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX DR
XX PR
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS
XX PS Claim 1; SEQ ID 1047; 2537pp + CD ROM; English.
XX CC
XX CC The present invention describes primer sets for synthesising 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesising polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ
XX SQ Sequence 855 BP; 257 A; 214 C; 176 G; 205 T; 3 other;
Query Match 11.1%; Score 438; DB 22; Length 855;
Best Local Similarity 99.8%; Pred. No. 2.9e-204;
Matches 488; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATCGACGACACGACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGCGTAT 60
DB 231 ATCGAAGACACGACATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGCGTAT 290
QY 61 TTAGCTGAACACGACATCGGGAAACAATGAGAGGAGTCGAGCGGCGCTCTCCCAAC 120
DB 291 TTAGCTGAACACGACATCGGGAAACAATGAGAGGAGTCGAGCGGCGCTCTCCCAAC 350
QY 121 CTTTGCCATTTTCGGAGTCCCTTCTGGGGCGCTGAAGAGGCGGAGGCCCAAGATGACCTT 180
DB 351 CTTTGCCATTTTCGGAGTCCCTTCTGGGGCGCTGAAGAGGCGGAGGCCCAAGATGACCTT 410
QY 181 CCAGATCTTTTCAGCCCTTCTGAGCCCAAGAAATAGACGAAAGTGTCAATTTGGCAAGA 240
DB 411 CCAGATCTTTTCAGCCCTTCTGAGCCCAAGAAATAGACGAAAGTGTCAATTTGGCAAGA 470
```

```
QY 241 CTGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAACGACTT 300
DB 471 CTGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAAACGACTT 530
QY 301 TCTCCTGATCAGATGAACACTCACCTAATTTAAGTTTTGAGCCTAACTTCTGCCAGAT 360
DB 531 TCTCCTGATCAGATGAACACTCACCTAATTTAAGTTTTGAGCCTAACTTCTGCCAGAT 590
QY 361 AACCTCGAAGTCCCACCAGCTCTAAGAGAAAGCCCCCAGGAGGCAAAAAGGCCACAGTAT 420
DB 591 AACCTCGAAGTCCCACCAGCTCTAAGAGAAAGCCCCCAGGAGGCAAAAAGGCCACAGTAT 650
QY 421 TGTCTGAAAACCCAGTCCAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGGAAGAG 480
DB 651 TGTCTGAAAACCCAGTCCAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGGAAGAG 710
QY 481 CTATCCTCC 489
DB 711 CTATCCTCC 719
RESULT 6
ABA06731
ID ABA06731 standard; cDNA; 702 BP.
XX AC ABA06731;
XX DT 10-JAN-2002 (first entry)
XX DE Human cDNA SEQ ID NO: 397.
XX KW Human; gene therapy; neural disorder; immune system disorder;
XX KW muscular disorder; reproductive disorder; gastrointestinal disorder;
XX KW pulmonary disorder; cardiovascular disorder; renal disorder;
XX KW proliferative disorder; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200154474-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01349.
XX PR 31-JAN-2000; 2000US-179065P.
XX PR 04-FEB-2000; 2000US-180628P.
XX PR 24-FEB-2000; 2000US-184664P.
XX PR 02-MAR-2000; 2000US-186350P.
XX PR 16-MAR-2000; 2000US-189874P.
XX PR 17-MAR-2000; 2000US-190076P.
XX PR 18-APR-2000; 2000US-198123P.
XX PR 19-MAY-2000; 2000US-205515P.
XX PR 07-JUN-2000; 2000US-209467P.
XX PR 28-JUN-2000; 2000US-214886P.
XX PR 30-JUN-2000; 2000US-215135P.
XX PR 07-JUL-2000; 2000US-216647P.
XX PR 07-JUL-2000; 2000US-216880P.
XX PR 11-JUL-2000; 2000US-217487P.
XX PR 11-JUL-2000; 2000US-217496P.
XX PR 14-JUL-2000; 2000US-218290P.
XX PR 26-JUL-2000; 2000US-220963P.
XX PR 26-JUL-2000; 2000US-220964P.
XX PR 14-AUG-2000; 2000US-224518P.
XX PR 14-AUG-2000; 2000US-224519P.
XX PR 14-AUG-2000; 2000US-225213P.
XX PR 14-AUG-2000; 2000US-225214P.
XX PR 14-AUG-2000; 2000US-225266P.
XX PR 14-AUG-2000; 2000US-225267P.
XX PR 14-AUG-2000; 2000US-225268P.
XX PR 14-AUG-2000; 2000US-225270P.
XX PR 14-AUG-2000; 2000US-225447P.
XX PR 14-AUG-2000; 2000US-225757P.
```

PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.

PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249246P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476161/51.
DR P-PSDB; ABB10509.
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
PS Claim 1; SEQ ID NO: 397; 859pp + Sequence Listing; English.
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
SQ Sequence 702 BP; 179 A; 195 C; 187 G; 137 T; 4 other;

Query Match 8.8%; Score 348; DB 22; Length 702;
Best Local Similarity 99.5%; Pred. No. 4.4e-160;
Matches 638; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 3177 AGAAAGAGACAAAGCCCTACAGGAACGCTTTTCCGACCACATTTCTCGAGGCTCC 3236
Db 2 AGAAAGAGACAAAGCCCTACAGGAACGCTTTTCCGACCACATTTCTCGAGGCTCC 61

QY 3237 TGGGGATATGGTAGCTCATGAGGGGCGCTCTGTCGGCTGGACTGTAAGTGAGTGGTTT 3296
Db 62 TGGGGATATGGTAGCTCATGAGGGGCGCTCTGTCGGCTGGACTGTAAGTGAGTGGTTT 121

QY 3297 ACCGCCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTGCTACACAGATGCCTCCCA 3356
Db 122 A-CGCCCCGGAGCTGACATGGCTACTCAATGCCAACCTGTGCTACACAGATGCCTCCCA 180

QY	3357	CAAGATGCTGGTCAGGAGACCGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCG	3416
Db	181	CAAGATGCTGGTCAGGAGACCGAGTCCACTCTCTGCTCATTTGACCCACTCACTCAGCG	240
QY	3417	CGACGAGGACCTATTAAGTGCATCGCTACCAACAACCGGCGAGAAATTCCTTTAGTCT	3476
Db	241	CGACG-AGGACCTATTAAGTGCATCGCTACCAACAACCGGCGAGAAATTCCTTTAGTCT	299
QY	3477	GGAGCTCTCTGTAGTACCCAAAGAGTGAAGAAAGCAACCTGTGTATCCTGGAGAAACTACA	3536
Db	300	GGAGCTCTCTGTAGTACCCAAAGAGTGAAGAAAGCAACCTGTGTATCCTGGAGAAACTACA	359
QY	3537	GAATCGGGTGTCCCGAAGGCCACCCCGTGGAGACTGGAGTGGCGGTGATAGGCATGCC	3596
Db	360	GAATCGGGTGTCCCGAAGGCCACCCCGTGGAGACTGGAGTGGCGGTGATAGGCATGCC	419
QY	3597	CCACCTGTCTTACTGGAAGAAAGACAATGAGACCATCCCTTGCACCGAGAGAGGAT	3656
Db	420	CCACCTGTCTTACTGGAAGAAAGACAATGAGACCATCCCTTGCACCGAGAGAGGAT	479
QY	3657	CAGTATGCACCGACACACAACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATC	3716
Db	480	CAGTATGCACCGACACACAACAGGGTATGCTGCTTCTCATTCAGCCAGCCCAAGAAATC	539
QY	3717	AGACGGTGGATGTACACGTTGTTCAGCCAAAGATGAAGCCGGCATCGTGTGTCACACTGC	3776
Db	540	AGACGGTGGATGTACACGTTGTTCAGCCAAAGATGAAGCCGGCATCGTGTGTCACACTGC	599
QY	3777	CAGCGTGATATATAGCTCAGTGGCACCACATCAGATCCAC	3817
Db	600	CAGCGTGATATATAGCTCAGTGGCACCACATCAGATCCAC	640
RESULT 7			
AAS28841			
ID	AAS28841 standard; cDNA; 702 BP.		
AC	AAS28841;		
DT	07-NOV-2001 (first entry)		
DE	Human immunoglobulin encoding cDNA SEQ ID NO 87.		
XX	Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;		
KW	antisease therapy; gene therapy; neurological disorder; renal disorder;		
KW	cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;		
KW	reproductive disorder; immune system disorder; proliferative disorder;		
KW	muscular disorder.		
OS	Homo sapiens.		
XX	WO200155315-A2.		
PD	02-AUG-2001.		
XX	17-JAN-2001; 2001WO-0501326.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184564.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226868.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		
PR	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0234223.		
PR	21-SEP-2000; 2000US-0234274.		
PR	25-SEP-2000; 2000US-0234597.		
PR	25-SEP-2000; 2000US-0234998.		
PR	26-SEP-2000; 2000US-0235484.		
PR	27-SEP-2000; 2000US-0235834.		
PR	29-SEP-2000; 2000US-0236327.		
PR	29-SEP-2000; 2000US-0236327.		
PR	29-SEP-2000; 2000US-0236367.		
PR	29-SEP-2000; 2000US-0236368.		
PR	29-SEP-2000; 2000US-0236369.		
PR	29-SEP-2000; 2000US-0236370.		
PR	02-OCT-2000; 2000US-0236802.		
PR	02-OCT-2000; 2000US-0237037.		
PR	02-OCT-2000; 2000US-0237038.		
PR	02-OCT-2000; 2000US-0237039.		
PR	02-OCT-2000; 2000US-0237040.		
PR	13-OCT-2000; 2000US-0239935.		
PR	13-OCT-2000; 2000US-0239937.		
PR	20-OCT-2000; 2000US-0240960.		
PR	20-OCT-2000; 2000US-0241221.		
PR	20-OCT-2000; 2000US-0241785.		
PR	20-OCT-2000; 2000US-0241786.		
PR	20-OCT-2000; 2000US-0241787.		
PR	20-OCT-2000; 2000US-0241808.		
PR	20-OCT-2000; 2000US-0241809.		
PR	20-OCT-2000; 2000US-0241826.		
PR	01-NOV-2000; 2000US-0244617.		
PR	08-NOV-2000; 2000US-0246474.		

KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
XX W0200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
PI
XX
XX WPI: 2000-602362/57.
XX P-PSDB; AAB40875.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 1115; 5507pp; English.
XX
XX AACT74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiparasitic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticoagulant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 392 BP; 114 A; 101 C; 82 G; 95 T; 0 other;
XX

Query Match 8.08; Score 318; DB 21; Length 392;
Best Local Similarity 99.7%; Pred. No. 2.3e-145;
Matches 368; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2416 CAGTTTCAGCCCGCTGTGTGTCCTCCCAATTCCTCTCTACACCGGATTTCAGAAC 2475
|||||
Db 1 CAGTTTCAGCCCGCTGTGTGTCCTCCCAATTCCTCTCTCTACACCGGATTTCAGAAC 60
|||||

QY 2476 CCACTGGCTTTCCCTCAGCTCTGTTCCTCCTCTCTCCCTGCCATCCACCAAAATGCC 2535
|||||
Db 61 CCACTGGCTTTCCCTCAGCTCTGTTCCTCCTCTCTCCCTGCCATCCACCAAAATGCC 120
|||||

QY 2536 ATGRCGTGCTAGAAAGTGCACCATCCATCCAGGGATTAGCGAGAAATATACA 2595
|||||

Db 121 ATGGGGTGCTAGAAAGTGCACCATCCATCCATCCAGGGATTAGCGAGAAATATACA 180
QY 2596 AAGTCTCTCAACAGTGAATGATGATAACATTCGTGAACCTAAGAACGCGAGTATTCTGA 2655
|||||
Db 181 AAGTCTCTCAACAGTGAATGATGATAACATTCGTGAACCTAAGAACGCGAGTATTCTGA 240
QY 2656 GACTTGGGGAAAAAATAACTTTTCAGTGATGTCAGACCAACACGAGGAGTACAAAATT 2715
|||||
Db 241 GACTTGGGGAAAAAATAACTTTTCAGTGATGTCAGACCAACACGAGGAGTACAAAATT 300
QY 2716 TCAAGCTTTGAGCAGAGCGCTGATGAATGAATAGAGTTTCGTTGGACGCTACTCCTGTT 2775
|||||
Db 301 TCAAGCTTTGAGCAGAGCGCTGATGAATGAATAGAGTTTCGTTGGACGCTACTCCTGTT 360
QY 2776 GATGAATCA 2784
|||||
Db 361 GATGAATCA 369
|||||

RESULT 9
ABN25334
ID ABN25334 standard; cDNA; 200 BP.
XX
XX ABN25334;
AC
XX
XX 24-JUN-2002 (first entry)
DT
XX
XX Human ORFX polynucleotide sequence SEQ ID NO:19145.
DE
XX
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX
XX Homo sapiens.
XX W0200192523-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US10836.
XX
XX 30-MAY-2000; 2000US-206132P.
XX 29-AUG-2000; 2000US-228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach MD;
XX
XX WPI: 2002-106308/14.
XX P-PSDB; ABP09582.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders -
XX
XX Disclosure; SEQ ID 19145; 1037pp; English.
XX

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification). ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ

transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.

N.B. the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 200 BP; 46 A; 70 C; 37 G; 47 T; 0 other;

Query Match 2.5%; Score 99; DB 24; Length 200;
Best Local Similarity 100.0%; Pred. No. 7.4e-38;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2121 TAGCAGGTGAAGGTCCTTCATCACAGAGTTCAGCTTGGCCGCCGGAAGTATTTCTT 2180
|||||
Db 96 TAGCAGGTGAAGGTCCTTCATCACAGAGTTCAGCTTGGCCGCCGGAAGTATTTCTT 155
|||||

QY 2181 CCCTCCAGCAACACCCAGCCGACCACTGTGGCCCTTC 2219
|||||
Db 156 CCCTCCAGCAACACCCAGCCGACCACTGTGGCCCTTC 194
|||||

RESULT 10
AAH09723/C
ID AAH09723 standard; cDNA; 530 BP.
XX
AC AAH09723;
DT 26-JUN-2001 (first entry)
DE Human cDNA clone (3'-primer) SEQ ID NO:6558.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
PI
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
PS Claim 3; SEQ ID 6558; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 530 BP; 121 A; 127 C; 100 G; 173 T; 9 other;

Query Match 0.6%; Score 22; DB 22; Length 530;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1462 CCTCGATCCATGGCAGAGCCAG 1483
|||||
Db 436 CCTCGATCCATGGCAGAGCCAG 415
|||||

RESULT 11
ABL75308
ID ABL75308 standard; cDNA; 281 BP.
XX
AC ABL75308;
DT 14-MAY-2002 (first entry)
XX
DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO:4682.
XX
KW Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;
KW inheritance; characteristic; growth; development; disease resistance;
KW environmental adaptability; quality; yield; molecular marker;
KW multigene trait; plant breeding; corn tassel; gene; ss.
XX
XX Zea mays.
XX
XX US2001051335-A1.
XX
XX 13-DEC-2001.
XX
XX 16-APR-1999; 99US-0294093.
XX
XX 21-APR-1998; 98US-082567P.
XX
XX (LALG/) LALGUDI R V.
PA (ITOL/) ITO L Y.
PA (SHER/) SHERMAN B K.
XX
XX Lalgudi RV, Ito LY, Sherman BK;
XX
XX WPI; 2002-163647/21.
DR
XX
XX Novel purified corn tassel-derived polynucleotide useful for
PT determining altered gene expression, to recover regulatory elements and
PT to follow inheritance of desirable characteristics through hybrid
PT breeding programs -
XX
XX Claim 1; SEQ ID 4682; 201pp; English.
PS
XX The present sequence describes a purified corn tassel-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence

CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (I) are also useful
 CC in the evaluation, and alteration of desired characteristics associated
 CC with growth and development, disease resistance, environmental
 CC adaptability, quality and yield, and as molecular markers for studying
 CC inheritance of multigene traits in a plant breeding program. (I) can be
 CC used to produce a tassel-specific profile of gene transcription, a
 CC transcript image, to clone regulatory elements for use in transformation
 CC vectors, to express a polypeptide, to identify, isolate or extend
 CC identical or related corn tassel nucleic acid sequences from DNA
 CC libraries, in nucleic acid hybridisation or amplification technologies,
 CC as query sequences to determine homology of known sequences, as probe
 CC for use in Southern or Northern hybridisation, and to identify the
 CC presence of and/or to determine the degree of similarity between two
 CC (or more) nucleic acid sequences.

XX
 SQ Sequence 281 BP; 54 A; 83 C; 91 G; 52 T; 1 other;

Query Match 0.5%; Score 21; DB 24; Length 281;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 GGAGCGGAGGCCAAGATGAC 177
 ||||||||||||||||
 Db 175 GGAGCGGAGGCCAAGATGAC 195

RESULT 12
 AAZ13687/C
 ID AAZ13687 standard; cDNA; 300 BP.
 AC AAZ13687;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:1156.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI; 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX
 PS Claim 1; Page 924-925; 2479pp; English.

XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX
 SQ Sequence 300 BP; 99 A; 40 C; 63 G; 98 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 300;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2791 GAAATTCACATGATGAGAT 2810
 ||||||||||||||||
 Db 199 GAAATTCACATGATGAGAT 180

RESULT 13
 AAZ09013
 ID AAZ09013 standard; cDNA; 399 BP.
 AC AAZ09013;
 XX
 DT 19-OCT-1999 (first entry)
 XX
 DE Differentiation Induction Subtraction Hybridization (DISH)-75 sequence.
 XX
 KW DAP; differentiation-associated protein; terminal differentiation;
 KW growth arrest; differentiation induction subtraction hybridization;
 KW DISH; melanoma; breast; lung; colorectal; prostate; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9937774-A2.
 XX
 PD 29-JUL-1999.
 XX
 PF 25-JAN-1999; 99WO-US01549.
 XX
 PR 29-MAY-1998; 98US-0087167.
 PR 26-JAN-1998; 98US-0073298.
 PR 11-FEB-1998; 98US-0074441.
 PR 12-MAR-1998; 98US-0077804.
 PR 25-MAR-1998; 98US-0079326.
 PR 28-APR-1998; 98US-0083195.
 PR 15-MAY-1998; 98US-0085609.
 PR 26-MAY-1998; 98US-0086829.
 XX
 PA (GENQ-) GENQUEST INC.
 XX
 PI Fisher PB, Huang F;
 XX
 DR WPI; 1999-479051/40.
 XX
 PT Differentiation-associated proteins and related polynucleotides,
 PT useful for vaccine and pharmaceuticals to inhibit cell growth

XX Claim 1; Fig 8; 144pp; English.

XX Sequences AA209006-209075 are Differentiation Induction Subtraction

CC Hybridization (DISH) sequences, which encode Differentiation-Associated

CC Proteins (DAPs). DAPs are associated with terminal differentiation and

CC growth arrest and the sequences encoding them range from 97-903 base

CC pairs in length. A DAP, a DAP fragment or a DAP polynucleotide may be

CC useful in inhibiting the development of cancer including prostate,

CC breast, lung and colorectal cancer, melanoma, astrocytoma or

CC glioblastoma multiforme. Determining the level of a DAP or its coding

CC sequence, in a tumour sample can be used to determine whether the tumour

CC is malignant. The progression of cancer can be monitored by measuring DAP

CC expression or activity levels over a period of time. An agent that

CC increases expression of a DAP can also be used to inhibit the development

CC of cancer.

XX SQ, Sequence 399 BP; 98 A; 91 C; 88 G; 122 T; 0 other;

Query Match 0.5%; Score 20; DB 20; Length 399;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 CAAAAAAGTATTTTAAATA 457

DB 182 CAAAAAAGTATTTTAAATA 201

|||||

RESULT 14

ABQ57885

ID ABQ57885 standard; cDNA; 551 BP.

XX AC ABQ57885;

XX OS Homo sapiens.

DT 02-AUG-2002 (first entry)

XX Human colon cancer related nucleotide sequence SEQ ID NO:1580.

DE Human; colon cancer; cancer; tissue profiling; forensic; mapping;

KW genetic analysis; diagnostic; antisense therapy; gene; ss.

XX WO200229086-A2.

PN 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30732.

XX 02-OCT-2000; 2000US-237271P.

XX (FARB) BAYER CORP.

PA Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;

PI Thiagalingam A, Lewis ME;

XX WPI; 2002-426115/45.

XX New isolated nucleic acid that is differentially expressed in cancer

PT tissues useful for determining the presence of colon cancer in a cell

PT or tissue type, and in antisense therapy -

XX Claim 1; Fig 1; 796pp; English.

XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially

CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins

CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be

CC used in antisense therapy. An antibody immunoreactive with a polypeptide

CC encoded by (I) is useful for detecting cancer in a patient sample, and

CC for detecting the presence or absence of a polynucleotide encoded by a

CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived

CC from (I) can be used for determining the presence of a nucleic acid which

CC hybridises to (I), and for determining the phenotype of cells in a sample

CC of cells from a patient. (I) is useful for determining the presence of

CC colon cancer in a cell or tissue type, for determining the presence or

CC state of other type of cancer, in antisense therapy, to generate

CC macroarrays on a solid surface, to identify a chromosome on which the

CC corresponding gene resides, and in tissue profiling, forensics, genetic

CC analysis, mapping and diagnostic applications. (I) can be used to raise

XX antibodies, and to screen for peptide analogues and antagonists.

SQ Sequence 551 BP; 124 A; 126 C; 105 G; 194 T; 2 other;

Query Match 0.5%; Score 20; DB 24; Length 551;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 CAAAAAAGTATTTTAAATA 457

DB 324 CAAAAAAGTATTTTAAATA 343

|||||

RESULT 15

AAZ15038/c

ID AAZ15038 standard; cDNA; 733 BP.

XX AC AAZ15038;

XX OS Homo sapiens.

DT 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:2507.

DE Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX WO9938972-A2.

PN 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

PT differentially expressed in different cell types

PT Claim 1; Page 1225; 2479pp; English.

XX The present invention describes a library of human polynucleotides

CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is

CC a method of detecting differentially expressed genes correlated with the

CC cancerous state of a mammalian cell, comprising detecting at least one

CC differentially expressed gene product in a test sample from a cell

CC suspected of being cancerous, where the gene product is encoded by one

CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The

CC polynucleotides can be used as a source of primers and probes, which can

CC be used for a variety of purpose, e.g. detection of expression levels,

CC mapping, tissue typing or profiling, forensics, genetic analysis and

CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists

Sequence 733 BP; 225 A; 142 C; 154 G; 205 T; 7 other;

Query Match	0.5%	Score 20;	DB 20;	Length 733;
Best Local Similarity	100.0%	Pred. No. 44;		
Matches 20;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	

Qy 2791 GAAATTCAACATGATGAGAT 2810
 |||||
 Db 259 GAAATTCAACATGATGAGAT 240

Search completed: November 30, 2002, 14:15:10
Job time : 572 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 30, 2002, 14:03:54 ; Search time 95 Seconds
(without alignments)
12793.263 Million cell updates/sec

Title: US-09-818-990B-1
Perfect score: 3963
Sequence: 1 atgcaagacgacgacataga.....tggagagtgaacttta 3963

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 1533381 residues

Word size : 0
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_NA: *
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2.6/ptodata/1/ina/PTUS_COMB.seq: *
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	19	0.5	360	4	US-09-118-554-17	Sequence 17, Appl
2	19	0.5	360	4	US-09-118-627-17	Sequence 17, Appl
3	19	0.5	360	4	US-09-602-877A-17	Sequence 2, Appl
c 4	19	0.5	1611	3	US-08-909-742-2	Sequence 2, Appl
c 5	19	0.5	1611	4	US-09-412-289-2	Sequence 3, Appl
6	19	0.5	2135	1	US-07-938-333A-3	Sequence 3, Appl
7	19	0.5	2135	1	US-08-660-216A-3	Sequence 3, Appl
8	19	0.5	6314	1	US-08-211-430-1	Sequence 1, Appl
9	18	0.5	861	1	US-07-988-273-1	Sequence 1, Appl
10	18	0.5	861	5	PCT-US93-12019-1	Sequence 1, Appl
11	18	0.5	1287	4	US-09-054-680-3	Sequence 3, Appl
12	18	0.5	2760	2	US-08-743-637B-20	Sequence 20, Appl
13	18	0.5	2760	3	US-08-526-840B-20	Sequence 20, Appl
14	18	0.5	2760	3	US-08-748-170A-3	Sequence 3, Appl
15	18	0.5	2760	3	US-09-047-148-1	Sequence 1, Appl
16	18	0.5	3451	2	US-08-743-637B-18	Sequence 18, Appl
17	18	0.5	3451	3	US-08-526-840B-18	Sequence 18, Appl
18	18	0.5	3865	4	US-09-149-476-296	Sequence 296, Appl
19	18	0.5	3984	4	US-09-442-100-3	Sequence 3, Appl
20	18	0.5	4843	3	US-08-986-485-1	Sequence 1, Appl
c 21	17	0.4	20	3	US-09-366-257-27	Sequence 27, Appl
22	17	0.4	291	1	US-07-922-723A-7	Sequence 7, Appl
23	17	0.4	291	1	US-07-799-828C-7	Sequence 7, Appl
24	17	0.4	291	1	US-08-074-275-7	Sequence 7, Appl
25	17	0.4	291	1	US-08-480-366-7	Sequence 7, Appl
26	17	0.4	291	2	US-07-952-277A-7	Sequence 7, Appl
27	17	0.4	390	4	US-09-254-465A-3	Sequence 3, Appl

507	0.4	17	4	US-09-328-111-826	Sequence 826, Appl
726	0.4	17	4	US-09-254-465A-4	Sequence 4, Appl
932	0.4	17	4	US-09-257-179-20	Sequence 20, Appl
1009	0.4	17	4	US-09-462-270-3	Sequence 3, Appl
1140	0.4	17	4	US-09-462-270-1	Sequence 1, Appl
1256	0.4	17	4	US-08-176-427B-9	Sequence 9, Appl
1256	0.4	17	4	US-08-356-060A-5	Sequence 5, Appl
1256	0.4	17	4	US-08-460-900C-5	Sequence 5, Appl
1256	0.4	17	4	US-08-674-509B-5	Sequence 5, Appl
1256	0.4	17	4	US-08-954-698-5	Sequence 5, Appl
1256	0.4	17	4	US-08-957-874-5	Sequence 5, Appl
1256	0.4	17	4	US-09-325-256-9	Sequence 9, Appl
1287	0.4	17	4	US-09-117-121-31	Sequence 31, Appl
1391	0.4	17	4	US-09-724-864-19	Sequence 19, Appl
1421	0.4	17	4	US-09-188-930-70	Sequence 70, Appl
1421	0.4	17	4	US-09-188-930-254	Sequence 254, Appl
1480	0.4	17	4	US-08-448-736-11	Sequence 11, Appl
1480	0.4	17	4	US-08-452-779-11	Sequence 11, Appl
1480	0.4	17	4	US-08-445-065-11	Sequence 11, Appl
1480	0.4	17	4	US-08-959-524-11	Sequence 11, Appl
1498	0.4	17	4	US-08-321-162-1	Sequence 1, Appl
1498	0.4	17	4	US-08-441-216-1	Sequence 1, Appl
1509	0.4	17	4	US-09-411-722-3	Sequence 3, Appl
1829	0.4	17	4	US-08-333-901-3	Sequence 3, Appl
1829	0.4	17	4	US-08-456-582-3	Sequence 3, Appl
1829	0.4	17	4	US-08-898-789-3	Sequence 3, Appl
1956	0.4	17	4	US-08-559-896B-1	Sequence 1, Appl
2055	0.4	17	4	US-09-197-218-1	Sequence 1, Appl
2139	0.4	17	4	US-08-611-280-3	Sequence 3, Appl
2139	0.4	17	4	US-09-195-940-3	Sequence 3, Appl
2139	0.4	17	4	US-09-562-466-3	Sequence 3, Appl
2165	0.4	17	4	US-09-366-257-1	Sequence 1, Appl
2276	0.4	17	4	US-09-172-841-56	Sequence 56, Appl
2288	0.4	17	4	US-09-135-232-1	Sequence 1, Appl
2362	0.4	17	4	US-09-232-200-48	Sequence 48, Appl
2362	0.4	17	4	US-09-232-201-48	Sequence 48, Appl
2362	0.4	17	4	US-08-628-434-3	Sequence 3, Appl
2638	0.4	17	4	US-08-306-691B-46	Sequence 46, Appl
2851	0.4	17	4	US-08-961-527-216	Sequence 216, Appl
2769	0.4	17	4	US-09-118-408-1	Sequence 1, Appl
2769	0.4	17	4	US-09-506-855-1	Sequence 1, Appl
2808	0.4	17	4	US-08-732-429-1	Sequence 1, Appl
2808	0.4	17	4	US-09-798-267-1	Sequence 1, Appl
2808	0.4	17	4	PCT-US95-05518-1	Sequence 1, Appl
2808	0.4	17	4	US-08-870-126-10	Sequence 10, Appl
3226	0.4	17	4	US-09-445-247-10	Sequence 10, Appl
3276	0.4	17	4	US-09-149-476-298	Sequence 298, Appl
3376	0.4	17	4	US-09-581-617-17	Sequence 17, Appl
3460	0.4	17	4	US-08-751-305-1	Sequence 1, Appl
4148	0.4	17	4	US-09-399-913-54	Sequence 54, Appl
5183	0.4	17	4	US-08-459-568-3	Sequence 3, Appl
5183	0.4	17	4	US-08-399-411-3	Sequence 3, Appl
5228	0.4	17	4	US-09-428-711A-15	Sequence 15, Appl
5804	0.4	17	4	US-09-369-364A-12	Sequence 12, Appl
5868	0.4	17	4	US-08-516-859A-3	Sequence 3, Appl
5868	0.4	17	4	US-09-586-472-3	Sequence 3, Appl
5868	0.4	17	4	US-09-528-706-3	Sequence 3, Appl
6596	0.4	17	4	US-09-575-602-11	Sequence 11, Appl
10523	0.4	17	4	US-09-453-702B-13	Sequence 13, Appl
15297	0.4	17	4	US-09-817-180-3	Sequence 3, Appl
19718	0.4	17	4	US-08-961-527-99	Sequence 99, Appl
20598	0.4	17	4	US-09-593-995-10	Sequence 10, Appl
59065	0.4	17	4	US-09-813-817-3	Sequence 3, Appl
59065	0.4	17	4	US-09-978-197-3	Sequence 3, Appl
80246	0.4	17	4	US-09-078-294-4	Sequence 4, Appl
80595	0.4	17	4	US-09-078-294-3	Sequence 3, Appl
99500	0.4	17	4	US-09-798-096-10	Sequence 10, Appl
20	0.4	16	4	US-09-158-863C-65	Sequence 65, Appl
20	0.4	16	4	US-09-254-465A-14	Sequence 14, Appl
24	0.4	16	4	US-09-158-863C-53	Sequence 53, Appl
24	0.4	16	4	PCT-US94-05085A-5	Sequence 5, Appl

c 101	16	0.4	24	5	PCT-US94-05085-5	Sequence 5, Appl	174	16	0.4	810	4	US-09-072-993C-5	Sequence 5, Appl
c 102	16	0.4	51	3	US-08-721-979A-67	Sequence 67, Appl	175	16	0.4	822	2	US-09-485-885-3	Sequence 3, Appl
c 103	16	0.4	51	3	US-08-836-501-67	Sequence 67, Appl	176	16	0.4	832	2	US-08-686-417-1	Sequence 1, Appl
c 104	16	0.4	51	4	US-09-654-289-67	Sequence 67, Appl	177	16	0.4	842	3	US-08-983-409-1	Sequence 1, Appl
105	16	0.4	52	3	US-08-602-145-6	Sequence 6, Appl	178	16	0.4	879	3	US-09-485-885-9	Sequence 9, Appl
106	16	0.4	83	1	US-08-657-012-10	Sequence 10, Appl	179	16	0.4	976	3	US-08-793-666-2	Sequence 2, Appl
107	16	0.4	83	3	US-09-013-872-10	Sequence 10, Appl	180	16	0.4	992	3	US-08-793-666-13	Sequence 12, Appl
108	16	0.4	83	4	US-09-184-198-10	Sequence 10, Appl	181	16	0.4	995	3	US-08-793-666-12	Sequence 13, Appl
109	16	0.4	83	4	US-09-633-653-10	Sequence 10, Appl	182	16	0.4	1000	1	US-07-965-274-3	Sequence 3, Appl
c 110	16	0.4	90	6	5219569-3	Patent No. 5219569	c 183	16	0.4	1056	4	US-09-067-782A-1	Sequence 1, Appl
c 111	16	0.4	99	3	US-08-721-979A-59	Sequence 59, Appl	c 184	16	0.4	1095	4	US-09-276-531-55	Sequence 55, Appl
c 112	16	0.4	99	3	US-08-836-501-59	Sequence 59, Appl	185	16	0.4	1116	4	US-09-485-885-5	Sequence 5, Appl
c 113	16	0.4	99	4	US-09-654-289-59	Sequence 59, Appl	186	16	0.4	1173	2	US-08-468-819-88	Sequence 88, Appl
c 114	16	0.4	153	3	US-08-721-979A-58	Sequence 58, Appl	187	16	0.4	1173	4	US-09-485-885-13	Sequence 13, Appl
c 115	16	0.4	153	3	US-08-836-501-58	Sequence 58, Appl	188	16	0.4	1182	4	US-09-336-536-38	Sequence 38, Appl
c 116	16	0.4	153	4	US-09-654-289-58	Sequence 58, Appl	c 189	16	0.4	1206	1	US-08-318-947A-1	Sequence 1, Appl
c 117	16	0.4	159	3	US-08-721-979A-57	Sequence 57, Appl	c 190	16	0.4	1206	4	US-08-795-303-1	Sequence 1, Appl
c 118	16	0.4	159	3	US-08-836-501-57	Sequence 57, Appl	191	16	0.4	1236	3	US-09-081-320-37	Sequence 37, Appl
c 119	16	0.4	159	4	US-09-654-289-57	Sequence 57, Appl	192	16	0.4	1236	4	US-09-574-141A-37	Sequence 37, Appl
c 120	16	0.4	165	3	US-08-721-979A-56	Sequence 56, Appl	193	16	0.4	1236	4	US-09-707-780-37	Sequence 37, Appl
c 121	16	0.4	165	3	US-08-836-501-56	Sequence 56, Appl	194	16	0.4	1237	1	US-08-197-792-40	Sequence 40, Appl
c 122	16	0.4	165	4	US-09-654-289-56	Sequence 56, Appl	195	16	0.4	1237	1	US-08-459-850-40	Sequence 40, Appl
c 123	16	0.4	171	3	US-08-721-979A-55	Sequence 55, Appl	196	16	0.4	1237	1	US-08-459-214-40	Sequence 40, Appl
c 124	16	0.4	171	3	US-08-836-501-55	Sequence 55, Appl	197	16	0.4	1240	1	US-08-240-372-2	Sequence 2, Appl
c 125	16	0.4	171	4	US-09-654-289-55	Sequence 55, Appl	198	16	0.4	1275	3	US-08-793-666-9	Sequence 9, Appl
c 126	16	0.4	177	3	US-08-721-979A-54	Sequence 54, Appl	199	16	0.4	1288	4	US-09-724-864-16	Sequence 16, Appl
c 127	16	0.4	177	3	US-08-836-501-54	Sequence 54, Appl	200	16	0.4	1353	4	US-08-913-014A-6	Sequence 6, Appl
c 128	16	0.4	177	4	US-09-654-289-54	Sequence 54, Appl	201	16	0.4	1358	1	US-08-181-271A-20	Sequence 20, Appl
c 129	16	0.4	183	3	US-08-721-979A-53	Sequence 53, Appl	202	16	0.4	1358	1	US-08-449-315-20	Sequence 20, Appl
c 130	16	0.4	183	3	US-08-836-501-53	Sequence 53, Appl	203	16	0.4	1358	1	US-08-444-803-20	Sequence 20, Appl
c 131	16	0.4	183	4	US-09-654-289-53	Sequence 53, Appl	204	16	0.4	1358	1	US-08-449-043-20	Sequence 20, Appl
c 132	16	0.4	189	5	PCT-US94-04026-8	Sequence 8, Appl	205	16	0.4	1358	1	US-08-456-265A-20	Sequence 20, Appl
c 133	16	0.4	303	3	US-08-721-979A-51	Sequence 51, Appl	206	16	0.4	1358	1	US-08-455-416-20	Sequence 20, Appl
c 134	16	0.4	303	3	US-08-836-501-51	Sequence 51, Appl	207	16	0.4	1358	1	US-08-455-244-20	Sequence 20, Appl
c 135	16	0.4	303	4	US-09-654-289-51	Sequence 51, Appl	208	16	0.4	1358	1	US-08-454-876-20	Sequence 20, Appl
136	16	0.4	374	4	US-09-228-986-14	Sequence 14, Appl	209	16	0.4	1358	2	US-08-457-364-20	Sequence 20, Appl
c 137	16	0.4	421	3	US-08-983-409-2	Sequence 2, Appl	210	16	0.4	1358	2	US-08-456-262-20	Sequence 20, Appl
c 138	16	0.4	434	4	US-09-370-838-248	Sequence 248, App	211	16	0.4	1358	2	US-08-456-240-20	Sequence 20, Appl
c 139	16	0.4	440	4	US-09-397-787-277	Sequence 277, App	212	16	0.4	1358	2	US-08-455-736-20	Sequence 20, Appl
140	16	0.4	453	4	US-09-134-001C-2664	Sequence 2664, Ap	213	16	0.4	1400	4	US-08-971-217-20	Sequence 20, Appl
141	16	0.4	456	4	US-09-056-105-29	Sequence 29, Appl	214	16	0.4	1400	4	US-09-350-600-20	Sequence 20, Appl
142	16	0.4	456	4	US-09-430-201-6	Sequence 6, Appl	215	16	0.4	1371	2	US-08-910-731-7	Sequence 7, Appl
c 143	16	0.4	468	4	US-09-067-782B-6	Sequence 6, Appl	216	16	0.4	1386	2	US-08-910-731-5	Sequence 5, Appl
c 144	16	0.4	483	4	US-09-370-838-141	Sequence 141, App	c 217	16	0.4	1392	4	US-09-163-444-1	Sequence 1, Appl
c 145	16	0.4	494	4	US-09-370-838-275	Sequence 275, App	c 218	16	0.4	1400	1	US-08-085-122-12	Sequence 12, Appl
146	16	0.4	519	3	US-08-860-165-13	Sequence 13, Appl	c 219	16	0.4	1400	4	US-08-442-108B-22	Sequence 22, Appl
147	16	0.4	519	4	US-09-359-382-13	Sequence 13, Appl	c 220	16	0.4	1481	1	US-08-136-922-1	Sequence 1, Appl
c 148	16	0.4	526	4	US-09-154-750A-71	Sequence 71, Appl	221	16	0.4	1491	4	US-08-913-014A-5	Sequence 5, Appl
149	16	0.4	536	1	US-08-329-704-1	Sequence 1, Appl	222	16	0.4	1561	1	US-07-968-971A-3	Sequence 3, Appl
c 150	16	0.4	536	2	US-08-472-604-1	Sequence 1, Appl	223	16	0.4	1561	1	US-07-824-247-44	Sequence 44, Appl
151	16	0.4	536	2	US-08-486-117-1	Sequence 1, Appl	224	16	0.4	1561	1	US-08-142-473A-4	Sequence 4, Appl
152	16	0.4	536	4	US-08-477-537-1	Sequence 1, Appl	225	16	0.4	1561	1	US-08-469-203A-4	Sequence 4, Appl
153	16	0.4	545	1	US-07-915-966C-1	Sequence 1, Appl	226	16	0.4	1561	1	US-08-469-203A-4	Sequence 4, Appl
154	16	0.4	545	2	US-08-771-182-1	Sequence 1, Appl	227	16	0.4	1561	3	US-08-470-204A-44	Sequence 44, Appl
155	16	0.4	545	3	US-08-853-194-1	Sequence 1, Appl	228	16	0.4	1561	4	US-08-926-522-10	Sequence 10, Appl
c 156	16	0.4	547	4	US-09-370-838-171	Sequence 171, App	c 229	16	0.4	1572	4	US-09-297-269-39	Sequence 39, Appl
157	16	0.4	570	5	PCT-US94-05085A-1	Sequence 1, Appl	230	16	0.4	1597	2	US-08-989-478-9	Sequence 9, Appl
158	16	0.4	570	5	PCT-US94-05085-1	Sequence 1, Appl	231	16	0.4	1597	3	US-08-986-685-9	Sequence 9, Appl
159	16	0.4	574	4	US-09-328-111-175	Sequence 175, App	232	16	0.4	1624	3	US-09-013-881-15	Sequence 15, Appl
c 160	16	0.4	585	4	US-09-404-879A-45	Sequence 45, Appl	c 233	16	0.4	1712	4	US-08-809-326A-8	Sequence 8, Appl
161	16	0.4	605	4	US-09-404-879A-95	Sequence 95, Appl	c 234	16	0.4	1735	4	US-09-163-444-3	Sequence 3, Appl
162	16	0.4	609	4	US-09-385-982-169	Sequence 169, App	c 235	16	0.4	1748	3	US-09-100-730-1	Sequence 1, Appl
163	16	0.4	636	4	US-09-404-879A-99	Sequence 99, Appl	236	16	0.4	1772	3	US-08-793-666-8	Sequence 8, Appl
c 164	16	0.4	697	4	US-09-404-879A-100	Sequence 100, App	237	16	0.4	1772	3	US-08-793-666-11	Sequence 11, Appl
c 165	16	0.4	720	4	US-08-913-014A-4	Sequence 4, Appl	c 238	16	0.4	1792	4	US-09-318-448-10	Sequence 10, Appl
c 166	16	0.4	750	4	US-09-307-973A-4	Sequence 4, Appl	239	16	0.4	1846	4	US-09-336-536-37	Sequence 37, Appl
c 167	16	0.4	751	4	US-09-227-357-117	Sequence 117, App	240	16	0.4	1896	4	US-09-343-011B-4	Sequence 4, Appl
c 168	16	0.4	776	1	US-08-216-233C-1	Sequence 1, Appl	c 241	16	0.4	1926	4	US-09-249-585A-2	Sequence 2, Appl
169	16	0.4	779	4	US-09-210-168-3	Sequence 3, Appl	c 242	16	0.4	1950	2	US-08-472-659-30	Sequence 30, Appl
170	16	0.4	790	1	US-08-117-083-7	Sequence 7, Appl	c 243	16	0.4	1950	2	US-08-474-661-30	Sequence 30, Appl
171	16	0.4	795	1	US-08-592-126-140	Sequence 140, App	c 244	16	0.4	1950	2	US-08-611-977-30	Sequence 30, Appl
172	16	0.4	801	3	US-08-860-165-9	Sequence 9, Appl	c 245	16	0.4	1954	1	US-08-097-829-3	Sequence 3, Appl
173	16	0.4	801	4	US-09-359-382-9	Sequence 9, Appl	c 246	16	0.4	1954	1	US-08-577-403-3	Sequence 3, Appl

c 247	16	0.4	1954	4	US-08-809-326A-13	Sequence 13, Appl	320	16	0.4	4066	3	US-09-215-966-21	Sequence 21, Appl
248	16	0.4	2011	2	US-08-989-478-6	Sequence 6, Appl	c 321	16	0.4	4081	4	US-08-999-774A-1	Sequence 1, Appl
249	16	0.4	2011	2	US-08-989-478-7	Sequence 7, Appl	322	16	0.4	4327	4	US-08-961-527-117	Sequence 117, App
250	16	0.4	2011	3	US-08-996-685-6	Sequence 6, Appl	323	16	0.4	4363	2	US-08-685-576-5	Sequence 5, Appl
251	16	0.4	2011	3	US-08-996-685-7	Sequence 7, Appl	324	16	0.4	4483	1	US-08-181-271A-5	Sequence 5, Appl
c 252	16	0.4	2012	4	US-08-020-143B-2	Sequence 2, Appl	325	16	0.4	4483	1	US-08-449-315-5	Sequence 5, Appl
c 253	16	0.4	2012	5	PCT-US92-10621-2	Sequence 2, Appl	326	16	0.4	4483	1	US-08-444-803-5	Sequence 5, Appl
c 254	16	0.4	2012	5	PCT-US94-02233-2	Sequence 2, Appl	327	16	0.4	4483	1	US-08-449-043-5	Sequence 5, Appl
255	16	0.4	2020	5	PCT-US95-13975-1	Sequence 1, Appl	328	16	0.4	4483	1	US-08-456-265A-5	Sequence 5, Appl
c 256	16	0.4	2058	4	US-09-560-639-4	Sequence 4, Appl	329	16	0.4	4483	1	US-08-455-416-5	Sequence 5, Appl
c 257	16	0.4	2119	2	US-08-381-691-17	Sequence 17, Appl	330	16	0.4	4483	1	US-08-455-244-5	Sequence 5, Appl
c 258	16	0.4	2137	4	US-08-991-862-1	Sequence 1, Appl	331	16	0.4	4483	2	US-08-454-876-5	Sequence 5, Appl
c 259	16	0.4	2214	1	US-07-985-458-1	Sequence 1, Appl	332	16	0.4	4483	2	US-08-457-364-5	Sequence 5, Appl
c 260	16	0.4	2217	1	US-08-712-241-4	Sequence 4, Appl	333	16	0.4	4483	2	US-08-456-262-5	Sequence 5, Appl
261	16	0.4	2253	6	5457037-2	Patent No. 5457037	334	16	0.4	4483	2	US-08-456-240-5	Sequence 5, Appl
c 262	16	0.4	2280	4	US-09-346-408-3	Sequence 3, Appl	335	16	0.4	4483	2	US-08-455-736-5	Sequence 5, Appl
c 263	16	0.4	2349	2	US-08-974-546-2	Sequence 2, Appl	336	16	0.4	4483	2	US-08-971-217-5	Sequence 5, Appl
264	16	0.4	2352	2	US-08-889-909A-21	Sequence 21, Appl	337	16	0.4	4483	4	US-09-350-600-5	Sequence 5, Appl
265	16	0.4	2352	4	US-09-156-163A-21	Sequence 21, Appl	338	16	0.4	4661	4	US-09-221-017B-970	Sequence 970, App
266	16	0.4	2421	3	US-08-851-843A-51	Sequence 51, Appl	339	16	0.4	4705	1	US-07-998-003A-96	Sequence 96, Appl
267	16	0.4	2421	4	US-08-974-549A-218	Sequence 218, App	340	16	0.4	4705	1	US-08-453-274B-96	Sequence 96, Appl
268	16	0.4	2421	4	US-08-854-050-51	Sequence 51, Appl	341	16	0.4	4705	1	US-08-453-695A-96	Sequence 96, Appl
269	16	0.4	2421	4	US-09-430-323-51	Sequence 51, Appl	342	16	0.4	4705	1	US-08-268-161A-96	Sequence 96, Appl
270	16	0.4	2429	4	US-09-386-493-3	Sequence 3, Appl	343	16	0.4	4705	2	US-08-453-702A-96	Sequence 96, Appl
271	16	0.4	2509	1	US-08-047-413-14	Sequence 14, Appl	344	16	0.4	4705	4	US-09-099-639-96	Sequence 96, Appl
272	16	0.4	2509	3	US-08-229-050-14	Sequence 14, Appl	345	16	0.4	4705	5	PCT-US93-12588-96	Sequence 96, Appl
273	16	0.4	2509	3	US-08-801-563-14	Sequence 14, Appl	346	16	0.4	4705	4	PCT-US95-08071-96	Sequence 96, Appl
c 274	16	0.4	2580	3	US-09-050-863-2	Sequence 2, Appl	347	16	0.4	5053	4	US-09-215-694-20	Sequence 20, Appl
c 275	16	0.4	2580	4	US-09-359-081-2	Sequence 2, Appl	c 347	16	0.4	5093	1	US-08-468-036-23	Sequence 23, Appl
276	16	0.4	2587	2	US-08-918-206-2	Sequence 2, Appl	348	16	0.4	5093	2	US-08-376-843-23	Sequence 23, Appl
277	16	0.4	2601	1	US-08-121-713D-53	Sequence 53, Appl	349	16	0.4	5093	2	US-08-130-114-1	Sequence 1, Appl
278	16	0.4	2601	1	US-08-835-268-53	Sequence 53, Appl	c 350	16	0.4	5262	4	US-08-520-373D-5	Sequence 5, Appl
279	16	0.4	2601	2	US-09-060-692-53	Sequence 53, Appl	351	16	0.4	5452	2	US-09-130-114-1	Sequence 1, Appl
280	16	0.4	2601	3	US-08-833-391-53	Sequence 53, Appl	352	16	0.4	5502	3	US-08-836-134-1	Sequence 1, Appl
281	16	0.4	2601	4	US-09-060-610-53	Sequence 53, Appl	353	16	0.4	5502	4	US-09-493-784-1	Sequence 1, Appl
282	16	0.4	2601	5	PCT-US94-10151A-53	Sequence 53, Appl	354	16	0.4	5655	2	US-08-989-478-1	Sequence 1, Appl
283	16	0.4	2625	6	5457037-4	Patent No. 5457037	355	16	0.4	5655	3	US-08-996-685-1	Sequence 1, Appl
c 284	16	0.4	2690	1	US-08-524-757-11	Sequence 11, Appl	356	16	0.4	5655	3	US-08-880-179-2	Sequence 2, Appl
c 285	16	0.4	2691	4	US-09-298-924-5	Sequence 5, Appl	c 357	16	0.4	5707	2	US-08-472-809B-8	Sequence 8, Appl
c 286	16	0.4	2710	1	US-08-487-748A-8	Sequence 8, Appl	358	16	0.4	6048	4	US-09-634-920-3	Sequence 3, Appl
c 287	16	0.4	2710	3	US-08-480-070C-8	Sequence 8, Appl	359	16	0.4	6176	3	US-08-911-321-6	Sequence 6, Appl
c 288	16	0.4	2710	3	US-08-829-525-8	Sequence 8, Appl	c 360	16	0.4	6345	2	US-08-472-809B-7	Sequence 7, Appl
c 289	16	0.4	2710	4	US-08-609-583A-8	Sequence 8, Appl	361	16	0.4	6462	4	US-09-792-594-10	Sequence 10, Appl
c 290	16	0.4	2710	4	US-08-937-399-8	Sequence 8, Appl	362	16	0.4	6485	3	US-09-081-320-2	Sequence 2, Appl
c 291	16	0.4	2710	4	US-09-310-367-8	Sequence 8, Appl	363	16	0.4	6485	4	US-09-574-141A-2	Sequence 2, Appl
c 292	16	0.4	2710	4	US-09-032-337-8	Sequence 8, Appl	364	16	0.4	6485	1	US-09-707-780-2	Sequence 2, Appl
293	16	0.4	2732	6	5217891-14	Patent No. 5217891	c 365	16	0.4	6756	4	US-08-151-574-31	Sequence 31, Appl
294	16	0.4	2733	2	US-08-846-021A-6	Sequence 6, Appl	c 366	16	0.4	6756	2	US-08-419-448-31	Sequence 31, Appl
c 295	16	0.4	2754	1	US-08-270-076A-10	Sequence 10, Appl	c 367	16	0.4	6756	4	US-09-233-510-31	Sequence 31, Appl
296	16	0.4	2809	3	US-08-448-194-5	Sequence 5, Appl	368	16	0.4	7198	4	US-09-453-702B-147	Sequence 147, App
297	16	0.4	2809	4	US-08-867-921-5	Sequence 5, Appl	c 369	16	0.4	7812	4	US-09-368-590-1	Sequence 1, Appl
298	16	0.4	2937	6	5208144-7	Patent No. 5208144	c 370	16	0.4	7886	2	US-08-751-189-2	Sequence 2, Appl
299	16	0.4	2964	2	US-08-286-819A-18	Sequence 18, Appl	c 371	16	0.4	7886	2	US-09-060-836-2	Sequence 2, Appl
300	16	0.4	2964	3	US-08-980-357-18	Sequence 18, Appl	c 372	16	0.4	7886	4	US-09-184-443-2	Sequence 2, Appl
c 301	16	0.4	3000	1	US-08-393-985-3	Sequence 3, Appl	373	16	0.4	7904	1	US-08-316-239B-1	Sequence 1, Appl
302	16	0.4	3048	2	US-08-313-200-2	Sequence 2, Appl	374	16	0.4	7904	1	US-08-316-239B-2	Sequence 2, Appl
303	16	0.4	3048	5	PCT-US93-03837-2	Sequence 2, Appl	375	16	0.4	7904	4	US-08-410-005-1	Sequence 1, Appl
c 304	16	0.4	3083	4	US-09-276-599-1	Sequence 1, Appl	376	16	0.4	7904	3	US-08-929-140-1	Sequence 1, Appl
c 305	16	0.4	3157	6	5198347-3	Patent No. 5198347	377	16	0.4	7904	4	US-09-560-579A-1	Sequence 1, Appl
306	16	0.4	3186	1	US-08-761-258-8	Sequence 8, Appl	378	16	0.4	8065	4	US-09-091-952A-6	Sequence 6, Appl
307	16	0.4	3186	2	US-08-977-306-8	Sequence 8, Appl	c 379	16	0.4	8575	5	PCT-US92-08258-6	Sequence 6, Appl
308	16	0.4	3190	2	US-08-286-819A-30	Sequence 30, Appl	380	16	0.4	8743	3	US-09-081-320-1	Sequence 1, Appl
309	16	0.4	3190	3	US-08-980-357-30	Sequence 30, Appl	381	16	0.4	8743	4	US-09-574-141A-1	Sequence 1, Appl
c 310	16	0.4	3267	4	US-08-257-963B-12	Sequence 12, Appl	382	16	0.4	8743	4	US-09-707-780-1	Sequence 1, Appl
c 311	16	0.4	3267	4	US-08-367-841A-12	Sequence 12, Appl	383	16	0.4	9069	4	US-08-961-527-97	Sequence 97, Appl
c 312	16	0.4	3267	5	PCT-US95-07201-12	Sequence 12, Appl	c 384	16	0.4	9600	4	US-08-910-647-1	Sequence 1, Appl
313	16	0.4	3283	4	US-09-651-656-16	Sequence 16, Appl	385	16	0.4	9600	4	US-09-620-925-1	Sequence 1, Appl
314	16	0.4	3283	4	US-09-650-855-16	Sequence 16, Appl	c 386	16	0.4	9737	4	US-09-479-122-22	Sequence 22, Appl
315	16	0.4	3286	2	US-08-363-124A-1	Sequence 1, Appl	c 387	16	0.4	9737	4	US-09-479-122-23	Sequence 23, Appl
316	16	0.4	3336	6	5457037-1	Patent No. 5457037	c 388	16	0.4	9737	4	US-09-479-122-28	Sequence 28, Appl
c 317	16	0.4	3501	1	US-08-524-757-5	Sequence 5, Appl	389	16	0.4	9793	1	US-08-470-202-56	Sequence 56, Appl
318	16	0.4	3748	2	US-08-686-417-2	Sequence 2, Appl	390	16	0.4	9793	1	US-08-471-770-56	Sequence 56, Appl
319	16	0.4	3972	4	US-09-770-170-1	Sequence 1, Appl	391	16	0.4	9793	2	US-08-468-059-56	Sequence 56, Appl
	16	0.4					392	16	0.4	9793	4	US-09-109-916-56	Sequence 56, Appl

c 539	15	0.4	384	4	US-09-134-001C-1146	Sequence 1146, Ap	612	15	0.4	627	4	US-09-710-092-7	Sequence 7, Appl
540	15	0.4	384	4	US-09-615-192A-232	Sequence 232, App	613	15	0.4	636	3	US-08-483-534A-1	Sequence 1, Appl
541	15	0.4	387	1	US-08-129-719-1	Sequence 1, Appl	614	15	0.4	648	4	US-09-328-111-258	Sequence 258, App
542	15	0.4	387	1	US-08-306-871-1	Sequence 1, Appl	615	15	0.4	651	4	US-09-198-119C-12	Sequence 12, Appl
543	15	0.4	387	1	US-08-569-959-1	Sequence 1, Appl	616	15	0.4	657	4	US-09-039-982A-38	Sequence 38, Appl
544	15	0.4	390	1	US-08-313-078A-51	Sequence 51, Appl	617	15	0.4	657	4	US-09-039-641-38	Sequence 38, Appl
c 545	15	0.4	400	1	US-08-208-008C-1	Sequence 1, Appl	618	15	0.4	657	4	US-09-039-762A-38	Sequence 38, Appl
546	15	0.4	424	1	US-08-229-781-30	Sequence 30, Appl	619	15	0.4	657	4	US-09-042-492D-38	Sequence 38, Appl
547	15	0.4	424	1	US-08-229-781-31	Sequence 31, Appl	620	15	0.4	657	4	US-08-913-612A-38	Sequence 38, Appl
548	15	0.4	424	1	US-08-630-918-30	Sequence 30, Appl	621	15	0.4	665	4	US-08-998-416-1045	Sequence 1045, Ap
549	15	0.4	424	1	US-08-630-918-31	Sequence 31, Appl	622	15	0.4	671	4	US-08-949-155-7	Sequence 7, Appl
550	15	0.4	424	4	US-09-004-422-30	Sequence 30, Appl	623	15	0.4	671	4	US-09-819-964-7	Sequence 7, Appl
551	15	0.4	424	4	US-09-004-422-31	Sequence 31, Appl	624	15	0.4	675	4	US-08-896-164-35	Sequence 35, Appl
552	15	0.4	428	4	US-09-004-731-37	Sequence 37, Appl	625	15	0.4	680	4	US-08-896-164-39	Sequence 39, Appl
c 553	15	0.4	428	4	US-09-004-731-39	Sequence 39, Appl	626	15	0.4	688	4	US-08-642-274D-45	Sequence 45, Appl
554	15	0.4	428	4	US-08-749-699-37	Sequence 37, Appl	627	15	0.4	688	4	US-08-952-014C-45	Sequence 45, Appl
c 555	15	0.4	428	4	US-08-749-699-39	Sequence 39, Appl	628	15	0.4	691	1	US-08-243-879A-35	Sequence 35, Appl
556	15	0.4	428	4	US-09-004-729-37	Sequence 37, Appl	629	15	0.4	691	1	US-08-243-879A-37	Sequence 37, Appl
c 557	15	0.4	428	4	US-09-004-729-39	Sequence 39, Appl	630	15	0.4	691	1	US-08-243-879A-39	Sequence 39, Appl
558	15	0.4	429	1	US-08-229-781-32	Sequence 32, Appl	631	15	0.4	691	1	US-08-243-879A-41	Sequence 41, Appl
559	15	0.4	429	1	US-08-630-918-32	Sequence 32, Appl	632	15	0.4	691	1	US-08-499-523-1	Sequence 1, Appl
560	15	0.4	429	4	US-09-004-422-32	Sequence 32, Appl	633	15	0.4	691	1	US-08-499-523-3	Sequence 3, Appl
561	15	0.4	434	3	US-08-866-340-25	Sequence 25, Appl	634	15	0.4	691	1	US-08-499-523-5	Sequence 5, Appl
562	15	0.4	434	4	US-09-103-875-31	Sequence 31, Appl	635	15	0.4	691	1	US-08-499-523-7	Sequence 7, Appl
c 563	15	0.4	436	4	US-09-397-787-257	Sequence 257, App	636	15	0.4	691	3	US-09-128-345-1	Sequence 1, Appl
564	15	0.4	436	4	US-09-397-787-288	Sequence 288, App	637	15	0.4	691	3	US-09-128-345-3	Sequence 3, Appl
c 565	15	0.4	440	4	US-09-397-787-311	Sequence 311, App	638	15	0.4	691	3	US-09-128-345-5	Sequence 5, Appl
c 566	15	0.4	444	3	US-08-688-988-44	Sequence 44, Appl	639	15	0.4	691	3	US-09-128-345-7	Sequence 7, Appl
c 567	15	0.4	458	4	US-09-060-758-311	Sequence 311, App	640	15	0.4	717	4	US-09-004-731-43	Sequence 43, Appl
c 568	15	0.4	461	4	US-09-385-958-163	Sequence 163, App	641	15	0.4	717	4	US-09-004-731-45	Sequence 45, Appl
569	15	0.4	474	3	US-08-928-799A-2	Sequence 2, Appl	642	15	0.4	717	4	US-08-749-699-43	Sequence 43, Appl
570	15	0.4	476	1	US-07-947-120-9	Sequence 9, Appl	643	15	0.4	717	4	US-08-749-699-45	Sequence 45, Appl
571	15	0.4	476	1	US-08-472-893A-9	Sequence 9, Appl	644	15	0.4	717	4	US-09-004-729-43	Sequence 43, Appl
572	15	0.4	476	3	US-08-947-492-9	Sequence 9, Appl	645	15	0.4	717	4	US-09-004-729-45	Sequence 45, Appl
573	15	0.4	498	2	US-08-810-572A-5	Sequence 5, Appl	646	15	0.4	721	4	US-08-998-416-606	Sequence 606, App
574	15	0.4	498	4	US-09-290-333-5	Sequence 5, Appl	647	15	0.4	723	4	US-08-466-465-3	Sequence 3, Appl
c 575	15	0.4	505	4	US-09-227-357-73	Sequence 73, Appl	648	15	0.4	724	4	US-09-535-008-40	Sequence 40, Appl
576	15	0.4	506	3	US-08-801-344-6	Sequence 6, Appl	649	15	0.4	726	4	US-09-032-215-49	Sequence 49, Appl
577	15	0.4	506	4	US-09-498-599-6	Sequence 6, Appl	650	15	0.4	726	4	US-09-032-215-50	Sequence 50, Appl
c 578	15	0.4	513	4	US-09-060-756-92	Sequence 92, Appl	651	15	0.4	726	4	US-09-039-982A-37	Sequence 37, Appl
579	15	0.4	534	4	US-09-132-316-17	Sequence 17, Appl	652	15	0.4	726	4	US-09-039-641-37	Sequence 37, Appl
580	15	0.4	537	4	US-09-397-787-48	Sequence 48, Appl	653	15	0.4	726	4	US-09-039-762A-37	Sequence 37, Appl
581	15	0.4	540	4	US-09-812-484-29	Sequence 29, Appl	654	15	0.4	726	4	US-09-042-492D-37	Sequence 37, Appl
582	15	0.4	545	4	US-08-991-789A-10	Sequence 10, Appl	655	15	0.4	726	4	US-08-913-612A-37	Sequence 37, Appl
583	15	0.4	545	4	US-09-062-451-10	Sequence 10, Appl	656	15	0.4	728	4	US-09-221-017B-10	Sequence 10, Appl
584	15	0.4	545	4	US-09-598-326-10	Sequence 10, Appl	657	15	0.4	737	4	US-08-998-416-932	Sequence 932, App
585	15	0.4	546	4	US-09-046-479-3	Sequence 3, Appl	658	15	0.4	750	4	US-08-998-416-257	Sequence 257, App
586	15	0.4	546	4	US-08-822-897C-3	Sequence 3, Appl	659	15	0.4	753	1	US-08-328-152A-16	Sequence 16, Appl
c 587	15	0.4	561	4	US-09-328-111-458	Sequence 458, App	660	15	0.4	753	4	US-08-466-465-1	Sequence 1, Appl
588	15	0.4	564	4	US-08-858-207A-187	Sequence 187, App	661	15	0.4	758	4	US-08-998-416-1017	Sequence 1017, Ap
589	15	0.4	584	4	US-09-812-484-12	Sequence 12, App	662	15	0.4	764	4	US-09-328-111-640	Sequence 640, App
c 590	15	0.4	604	4	US-08-998-416-243	Sequence 243, App	663	15	0.4	769	4	US-08-858-207A-154	Sequence 154, App
591	15	0.4	606	4	US-09-385-982-60	Sequence 60, Appl	664	15	0.4	774	4	US-09-166-350-5	Sequence 5, Appl
c 592	15	0.4	606	4	US-09-385-982-413	Sequence 413, App	665	15	0.4	789	4	US-09-164-193-6	Sequence 6, Appl
c 593	15	0.4	607	4	US-09-556-877-282	Sequence 282, App	666	15	0.4	789	4	US-09-221-448A-6	Sequence 6, Appl
c 594	15	0.4	607	4	US-09-620-413C-282	Sequence 282, App	667	15	0.4	792	4	US-09-134-001C-524	Sequence 524, App
c 595	15	0.4	609	4	US-09-328-111-417	Sequence 417, App	668	15	0.4	792	4	US-09-134-001C-2275	Sequence 2275, Ap
596	15	0.4	613	4	US-09-385-982-219	Sequence 219, App	669	15	0.4	804	4	US-09-134-001C-683	Sequence 683, App
c 597	15	0.4	618	4	US-09-385-982-218	Sequence 218, App	670	15	0.4	807	4	US-09-276-531-82	Sequence 82, Appl
598	15	0.4	627	1	US-08-181-271A-23	Sequence 23, Appl	671	15	0.4	836	4	US-09-352-990-7	Sequence 7, Appl
599	15	0.4	627	1	US-08-449-315-23	Sequence 23, Appl	672	15	0.4	841	4	US-09-004-731-40	Sequence 40, Appl
600	15	0.4	627	1	US-08-444-803-23	Sequence 23, Appl	673	15	0.4	841	4	US-09-004-731-42	Sequence 42, Appl
601	15	0.4	627	1	US-08-449-043-23	Sequence 23, Appl	674	15	0.4	841	4	US-09-032-215-46	Sequence 46, Appl
602	15	0.4	627	1	US-08-456-265A-23	Sequence 23, Appl	675	15	0.4	841	4	US-09-032-215-48	Sequence 48, Appl
603	15	0.4	627	1	US-08-455-416-23	Sequence 23, Appl	676	15	0.4	841	4	US-08-749-699-40	Sequence 40, Appl
604	15	0.4	627	1	US-08-455-244-23	Sequence 23, Appl	677	15	0.4	841	4	US-08-749-699-42	Sequence 42, Appl
605	15	0.4	627	1	US-08-454-876-23	Sequence 23, Appl	678	15	0.4	841	4	US-09-004-729-40	Sequence 40, Appl
606	15	0.4	627	2	US-08-457-364-23	Sequence 23, Appl	679	15	0.4	841	4	US-09-004-729-42	Sequence 42, Appl
607	15	0.4	627	2	US-08-456-262-23	Sequence 23, Appl	680	15	0.4	844	4	US-09-812-484-6	Sequence 6, Appl
608	15	0.4	627	2	US-08-456-240-23	Sequence 23, Appl	681	15	0.4	852	4	US-09-222-939-1	Sequence 1, Appl
609	15	0.4	627	2	US-08-455-736-23	Sequence 23, Appl	682	15	0.4	855	4	US-09-247-373B-37	Sequence 37, Appl
610	15	0.4	627	2	US-08-971-217-23	Sequence 23, Appl	683	15	0.4	855	6	5185441-40	Patent No. 5185441
611	15	0.4	627	4	US-09-350-600-23	Sequence 23, Appl	684	15	0.4	855	6	5223394-3	Patent No. 5223394

c 831	15	0.4	1244	5	PCT-US96-02331-12	Sequence 12, Appl	c 904	15	0.4	1505	5	PCT-US93-10301-5	Sequence 5, Appl
c 832	15	0.4	1248	2	US-08-391-916A-1	Sequence 1, Appl	c 905	15	0.4	1531	2	US-08-833-963C-1	Sequence 1, Appl
c 833	15	0.4	1257	4	US-08-791-115B-8	Sequence 8, Appl	c 906	15	0.4	1542	1	US-08-249-112-1	Sequence 1, Appl
c 834	15	0.4	1263	4	US-08-811-481-31	Sequence 31, Appl	c 907	15	0.4	1542	5	PCT-US95-06556-1	Sequence 1, Appl
c 835	15	0.4	1263	5	PCT-US96-10602-9	Sequence 9, Appl	c 908	15	0.4	1548	4	US-08-706-945D-46	Sequence 46, Appl
c 836	15	0.4	1265	4	US-09-964-850-3	Sequence 3, Appl	c 909	15	0.4	1561	3	US-08-965-850-1	Sequence 1, Appl
c 837	15	0.4	1269	4	US-09-710-092-11	Sequence 11, Appl	c 910	15	0.4	1565	3	US-08-468-846-1	Sequence 1, Appl
c 838	15	0.4	1269	5	PCT-US93-00893-4	Sequence 4, Appl	c 911	15	0.4	1565	4	US-08-915-096A-1	Sequence 1, Appl
c 839	15	0.4	1272	4	US-09-134-001C-2373	Sequence 2373, Ap	c 912	15	0.4	1579	4	US-09-227-357-59	Sequence 59, Appl
c 840	15	0.4	1284	4	US-09-134-001C-2304	Sequence 2304, Ap	c 913	15	0.4	1607	1	US-08-446-908-1	Sequence 1, Appl
c 841	15	0.4	1287	1	US-08-471-058-8	Sequence 8, Appl	c 914	15	0.4	1607	1	US-08-231-205A-1	Sequence 1, Appl
c 842	15	0.4	1287	2	US-08-944-530-3	Sequence 3, Appl	c 915	15	0.4	1607	2	US-08-871-161-1	Sequence 1, Appl
c 843	15	0.4	1287	3	US-08-471-057-8	Sequence 8, Appl	c 916	15	0.4	1621	2	US-08-933-750C-96	Sequence 96, Appl
c 844	15	0.4	1293	4	US-09-182-145-13	Sequence 13, Appl	c 917	15	0.4	1621	3	US-09-234-613-96	Sequence 96, Appl
c 845	15	0.4	1293	4	US-09-182-145-14	Sequence 14, Appl	c 918	15	0.4	1659	1	US-08-203-716-1	Sequence 1, Appl
c 846	15	0.4	1300	3	US-08-971-782-3	Sequence 3, Appl	c 919	15	0.4	1659	1	US-08-440-179-1	Sequence 1, Appl
c 847	15	0.4	1300	4	US-09-309-026-3	Sequence 3, Appl	c 920	15	0.4	1659	5	PCT-US91-02339-2	Sequence 2, Appl
c 848	15	0.4	1314	2	US-08-440-845D-9	Sequence 9, Appl	c 921	15	0.4	1659	5	PCT-US91-02339-2	Sequence 2, Appl
c 849	15	0.4	1314	3	US-08-868-458-9	Sequence 9, Appl	c 922	15	0.4	1661	4	US-09-149-476-246	Sequence 246, Appl
c 850	15	0.4	1317	2	US-08-464-517-34	Sequence 34, Appl	c 923	15	0.4	1675	1	US-07-688-352C-29	Sequence 29, Appl
c 851	15	0.4	1317	2	US-08-246-361A-34	Sequence 34, Appl	c 924	15	0.4	1675	2	US-08-474-379C-29	Sequence 29, Appl
c 852	15	0.4	1317	3	US-08-463-772-34	Sequence 34, Appl	c 925	15	0.4	1675	3	US-09-146-249A-29	Sequence 29, Appl
c 853	15	0.4	1317	5	PCT-US93-05000-34	Sequence 34, Appl	c 926	15	0.4	1675	3	US-08-206-189B-29	Sequence 29, Appl
c 854	15	0.4	1317	5	PCT-US93-05000-36	Sequence 36, Appl	c 927	15	0.4	1675	5	PCT-US91-02714-28	Sequence 28, Appl
c 855	15	0.4	1320	2	US-08-695-412B-13	Sequence 13, Appl	c 928	15	0.4	1683	2	US-08-824-405-5	Sequence 5, Appl
c 856	15	0.4	1320	4	US-09-255-154D-13	Sequence 13, Appl	c 929	15	0.4	1684	4	US-08-961-527-258	Sequence 258, Appl
c 857	15	0.4	1325	1	US-08-306-691B-51	Sequence 51, Appl	c 930	15	0.4	1689	4	US-09-311-924-1	Sequence 1, Appl
c 858	15	0.4	1325	2	US-08-464-517-1	Sequence 1, Appl	c 931	15	0.4	1691	3	US-09-013-881-12	Sequence 12, Appl
c 859	15	0.4	1325	2	US-08-246-361A-1	Sequence 1, Appl	c 932	15	0.4	1697	1	US-08-181-271A-104	Sequence 104, Appl
c 860	15	0.4	1325	3	US-08-463-772-1	Sequence 1, Appl	c 933	15	0.4	1697	1	US-08-449-315-104	Sequence 104, Appl
c 861	15	0.4	1325	5	PCT-US93-05000-1	Sequence 1, Appl	c 934	15	0.4	1697	1	US-08-444-803-104	Sequence 104, Appl
c 862	15	0.4	1336	1	US-08-107-684B-1	Sequence 1, Appl	c 935	15	0.4	1697	1	US-08-449-043-104	Sequence 104, Appl
c 863	15	0.4	1337	2	US-08-743-637B-32	Sequence 32, Appl	c 936	15	0.4	1697	1	US-08-456-265A-104	Sequence 104, Appl
c 864	15	0.4	1337	3	US-08-526-840B-32	Sequence 32, Appl	c 937	15	0.4	1697	1	US-08-455-416-104	Sequence 104, Appl
c 865	15	0.4	1363	4	US-09-280-116-55	Sequence 55, Appl	c 938	15	0.4	1697	1	US-08-455-244-104	Sequence 104, Appl
c 866	15	0.4	1368	3	US-09-142-469-1	Sequence 1, Appl	c 939	15	0.4	1697	2	US-08-454-876-104	Sequence 104, Appl
c 867	15	0.4	1370	1	US-08-592-125-79	Sequence 79, Appl	c 940	15	0.4	1697	2	US-08-457-364-104	Sequence 104, Appl
c 868	15	0.4	1373	2	US-08-394-189B-3	Sequence 3, Appl	c 941	15	0.4	1697	2	US-08-456-263-104	Sequence 104, Appl
c 869	15	0.4	1373	5	PCT-US93-05705-3	Sequence 3, Appl	c 942	15	0.4	1697	2	US-08-456-240-104	Sequence 104, Appl
c 870	15	0.4	1377	2	US-08-810-572A-1	Sequence 1, Appl	c 943	15	0.4	1697	2	US-08-455-736-104	Sequence 104, Appl
c 871	15	0.4	1377	4	US-09-290-533-20	Sequence 20, Appl	c 944	15	0.4	1697	3	US-08-971-211-104	Sequence 104, Appl
c 872	15	0.4	1378	1	US-08-075-533-20	Sequence 20, Appl	c 945	15	0.4	1697	3	US-09-188-930-264	Sequence 264, Appl
c 873	15	0.4	1378	2	US-08-948-176-20	Sequence 20, Appl	c 946	15	0.4	1697	4	US-09-350-600-104	Sequence 104, Appl
c 874	15	0.4	1378	5	PCT-US91-09160-20	Sequence 20, Appl	c 947	15	0.4	1706	1	US-07-906-871-11	Sequence 11, Appl
c 875	15	0.4	1385	4	US-08-858-207A-41	Sequence 41, Appl	c 948	15	0.4	1726	4	US-09-232-160-14	Sequence 14, Appl
c 876	15	0.4	1386	4	US-09-364-230-23	Sequence 23, Appl	c 949	15	0.4	1726	4	US-09-812-484-11	Sequence 11, Appl
c 877	15	0.4	1405	4	US-08-123-615-2	Sequence 1, Appl	c 950	15	0.4	1727	1	US-08-129-719-14	Sequence 14, Appl
c 878	15	0.4	1410	5	PCT-US95-15428-14	Sequence 14, Appl	c 951	15	0.4	1727	1	US-08-306-871-14	Sequence 14, Appl
c 879	15	0.4	1412	2	US-08-948-176-23	Sequence 23, Appl	c 952	15	0.4	1727	1	US-08-569-959-14	Sequence 14, Appl
c 880	15	0.4	1412	4	US-09-303-592-1	Sequence 1, Appl	c 953	15	0.4	1728	3	US-08-379-802-1	Sequence 1, Appl
c 881	15	0.4	1417	1	US-08-117-362-31	Sequence 31, Appl	c 954	15	0.4	1728	3	US-09-048-129-1	Sequence 1, Appl
c 882	15	0.4	1417	1	US-08-486-924-31	Sequence 31, Appl	c 955	15	0.4	1728	4	US-09-048-079-1	Sequence 1, Appl
c 883	15	0.4	1434	2	US-08-903-801-2	Sequence 2, Appl	c 956	15	0.4	1736	1	US-08-518-474-7	Sequence 7, Appl
c 884	15	0.4	1434	4	US-09-295-055-2	Sequence 2, Appl	c 957	15	0.4	1737	1	US-09-710-092-15	Sequence 15, Appl
c 885	15	0.4	1448	1	US-08-298-426-1	Sequence 1, Appl	c 958	15	0.4	1743	2	US-08-841-178-20	Sequence 20, Appl
c 886	15	0.4	1449	2	US-08-705-868-2	Sequence 2, Appl	c 959	15	0.4	1754	1	US-08-229-781-27	Sequence 27, Appl
c 887	15	0.4	1449	3	US-09-123-615-2	Sequence 2, Appl	c 960	15	0.4	1754	1	US-08-630-918-27	Sequence 27, Appl
c 888	15	0.4	1458	1	US-08-611-107-1	Sequence 1, Appl	c 961	15	0.4	1754	4	US-09-004-422-27	Sequence 27, Appl
c 889	15	0.4	1458	2	US-08-422-560A-1	Sequence 1, Appl	c 962	15	0.4	1765	1	US-08-336-778-2	Sequence 2, Appl
c 890	15	0.4	1458	4	US-08-468-793-1	Sequence 1, Appl	c 963	15	0.4	1766	2	US-08-453-848-8	Sequence 8, Appl
c 891	15	0.4	1460	4	US-09-149-476-87	Sequence 87, Appl	c 964	15	0.4	1766	4	US-09-169-027-8	Sequence 8, Appl
c 892	15	0.4	1472	1	US-08-123-161A-9	Sequence 9, Appl	c 965	15	0.4	1767	3	US-09-199-859-1	Sequence 1, Appl
c 893	15	0.4	1472	1	US-08-483-278-9	Sequence 9, Appl	c 966	15	0.4	1767	4	US-09-227-357-42	Sequence 42, Appl
c 894	15	0.4	1494	1	US-08-318-905-21	Sequence 21, Appl	c 967	15	0.4	1773	4	US-08-906-156A-11	Sequence 11, Appl
c 895	15	0.4	1494	1	US-08-483-232-21	Sequence 21, Appl	c 968	15	0.4	1776	1	US-08-298-426-3	Sequence 3, Appl
c 896	15	0.4	1494	1	US-08-483-140-21	Sequence 21, Appl	c 969	15	0.4	1777	1	US-08-229-781-54	Sequence 54, Appl
c 897	15	0.4	1494	2	US-08-485-938A-21	Sequence 21, Appl	c 970	15	0.4	1777	1	US-08-630-918-54	Sequence 54, Appl
c 898	15	0.4	1494	2	US-08-910-041-21	Sequence 21, Appl	c 971	15	0.4	1777	4	US-09-004-422-54	Sequence 54, Appl
c 899	15	0.4	1494	3	US-09-328-474-21	Sequence 21, Appl	c 972	15	0.4	1803	4	US-08-976-259-71	Sequence 71, Appl
c 900	15	0.4	1494	3	US-09-100-546-21	Sequence 21, Appl	c 973	15	0.4	1817	3	US-08-564-496C-14	Sequence 14, Appl
c 901	15	0.4	1494	3	US-09-010-715-21	Sequence 21, Appl	c 974	15	0.4	1820	4	US-09-732-199A-3	Sequence 3, Appl
c 902	15	0.4	1494	4	US-09-577-758-21	Sequence 21, Appl	c 975	15	0.4	1824	2	US-08-822-701-1	Sequence 1, Appl
c 903	15	0.4	1505	2	US-08-428-243-5	Sequence 5, Appl	c 976	15	0.4	1824	3	US-08-935-855-1	Sequence 1, Appl

```
c 977      15      0.4      1831      4      US-09-422-576D-1
978      15      0.4      1839      1      US-08-442-248-3
979      15      0.4      1839      1      US-08-440-815-3
980      15      0.4      1839      4      US-08-486-449-3
981      15      0.4      1841      2      US-08-820-521-1
982      15      0.4      1841      4      US-09-248-715-1
983      15      0.4      1841      4      US-09-248-715-1
984      15      0.4      1842      4      US-08-961-527-85
985      15      0.4      1843      1      US-08-774-169-2
986      15      0.4      1843      1      US-08-499-523-9
987      15      0.4      1843      3      US-09-128-345-9
988      15      0.4      1854      5      PCT-US94-01101-1
989      15      0.4      1859      4      US-09-286-529-19
990      15      0.4      1859      4      US-09-286-529-19
991      15      0.4      1869      4      US-09-406-045-2
992      15      0.4      1869      4      US-09-406-045-2
993      15      0.4      1880      1      US-08-247-475-49
994      15      0.4      1880      1      US-08-479-650-49
995      15      0.4      1880      1      US-08-191-866D-80
996      15      0.4      1880      1      US-08-674-169-49
997      15      0.4      1880      2      US-08-185-949B-80
998      15      0.4      1944      4      US-09-149-476-144
999      15      0.4      1962      4      US-08-791-115B-3
c1000     15      0.4      1964      4      US-09-434-613-2
```

ALIGNMENTS

```
RESULT 1
US-09-118-554-17
; Sequence 17, Application US/09118554A
; Patent No. 6365348
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS FOR DIAGNOSIS OF BREAST CANCER AND
; TITLE OF INVENTION: METHODS FOR THEIR USE
; FILE REFERENCE: 210121.450C1
; CURRENT APPLICATION NUMBER: US/09/118,554A
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 08/988,255
; EARLIER FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (56)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (81)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (85)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (89)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (132)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (146)
; OTHER INFORMATION: Wherein n is a, c, g or t
```

```
;
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (206)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (226)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (245)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (254)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (268)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (273)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (275)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (286)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (289)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (292)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (294)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (299)
; OTHER INFORMATION: Wherein n is a, c, g or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (347)
; OTHER INFORMATION: Wherein n is a, c, g or t
;
US-09-118-554-17
```

```
Query Match      0.5%; Score 19; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2275 AGCAAGAAAGCCTCTAG 2293
Db 35 AGCAAGAAAGCCTCTAG 53
|||||
```

```
RESULT 2
US-09-118-627-17
; Sequence 17, Application US/09118627A
; Patent No. 6379951
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
```

APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF BREAST CANCER
FILE REFERENCE: 210121.446C1
CURRENT APPLICATION NUMBER: US/09/118,627A
CURRENT FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 67
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 360
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(360)
OTHER INFORMATION: n = A,T,C or G
US-09-118-627-17

Query Match 0.5%; Score 19; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2275 AGCAAGAAGGCTCTTAG 2293

Db 35 AGCAAGAAGGCTCTTAG 53

RESULT 3

US-09-602-877A-17
Sequence 17, Application US/09602877A
Patent No. 6432707

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.446C5
CURRENT APPLICATION NUMBER: US/09/602,877A
CURRENT FILING DATE: 2000-06-22
NUMBER OF SEQ ID NOS: 107
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 360
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(360)
OTHER INFORMATION: n = A,T,C or G
US-09-602-877A-17

Query Match 0.5%; Score 19; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2275 AGCAAGAAGGCTCTTAG 2293

Db 35 AGCAAGAAGGCTCTTAG 53

RESULT 4

US-08-909-742-2/c
Sequence 2, Application US/08909742
Patent No. 6007991

GENERAL INFORMATION:
APPLICANT: Vimala S. Sivaraman
APPLICANT: Hsien-Yu Wang
APPLICANT: Craig C. Malbon
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES FOR MITOGEN-
TITLE OF INVENTION: ACTIVATED PROTEIN KINASES AS THERAPY FOR
BREAST CANCER
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann & Baron, LLP
STREET: 350 Jericho Turnpike
CITY: Jericho
STATE: New York
COUNTRY: USA
ZIP: 11753
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909,742
FILING DATE: August 12, 1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/831,994
FILING DATE: April 1, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,520
FILING DATE: March 28, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Adams, Lindsay S.
REGISTRATION NUMBER: 36,425
REFERENCE/DOCKET NUMBER: 178-225 CIP II
TELEPHONE: (516) 822-3550
TELEFAX: (516) 822-3582
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1611 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-909-742-2

Query Match 0.5%; Score 19; DB 3; Length 1611;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1499 TGGTCATTGCTGAGGTCTT 1517

Db 701 TGGTCATTGCTGAGGTCTT 683

RESULT 5

US-09-412-289-2/c
Sequence 2, Application US/09412289
Patent No. 6271210

GENERAL INFORMATION:
APPLICANT: Sivaraman, Vimala S.
APPLICANT: Wang, Hsien-Yu
APPLICANT: Malbon, Craig C.
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES FOR MITOGEN-ACTIVATED
TITLE OF INVENTION: PROTEIN KINASES AS THERAPY FOR CANCER
FILE REFERENCE: Seq. 1-4 (178-225 CIP II/CON)
CURRENT APPLICATION NUMBER: US/09/412,289
CURRENT FILING DATE: 1999-10-05

EARLIER APPLICATION NUMBER: 08/909,742
EARLIER FILING DATE: 1997-08-12
EARLIER APPLICATION NUMBER: 08/831,994
EARLIER FILING DATE: 1997-04-01
EARLIER APPLICATION NUMBER: 08/827,520
EARLIER FILING DATE: 1997-03-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1611
TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: blank
US-09-412-289-2

Query Match 0.5%; Score 19; DB 4; Length 1611;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1499 TGGTCATTGCTGAGGCTTT 1517
|||||
Db 701 TGGTCATTGCTGAGGCTTT 683

RESULT 6

US-07-938-333A-3
; Sequence 3, Application US/07938333A
; Patent No. 5439822
; GENERAL INFORMATION:
; APPLICANT: Katsumata, Ryoichi
; APPLICANT: Takano, Yutaka
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50inch
; COMPUTER: Toshiba J-3100 GT (IBM PC compatible)
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Apple Macintosh Hypercard 1.2.2,
; SOFTWARE: Apple File Exchange J1-1.1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/938,333A
; FILING DATE: 19920828
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lippert, Nels T.
; REGISTRATION NUMBER: 25,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-819-8200
; TELEFAX: 212-354-8113
; TELEX: 126201
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2135 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13032
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 514 to 1806
; IDENTIFICATION METHOD: E
US-07-938-333A-3

Query Match 0.5%; Score 19; DB 1; Length 2135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AAGTCCAGAAGGAAGCTCG 854
|||||
Db 1504 AAGTCCAGAAGGAAGCTCG 1522

RESULT 7

; TITLE OF INVENTION: NUCLEIC SEQUENCE OF THE GENE ASSOCIATED WITH
; TITLE OF INVENTION: X CHROMOSOME LINKED KALLMANN SYNDROME, CORRESPONDING

US-08-660-216A-3
; Sequence 3, Application US/08660216A
; Patent No. 5700661
; GENERAL INFORMATION:
; APPLICANT: Katsumata, Ryoichi
; APPLICANT: Takano, Yutaka
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50inch
; COMPUTER: Toshiba J-3100 GT (IBM PC compatible)
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Apple Macintosh Hypercard 1.2.2,
; SOFTWARE: Apple File Exchange J1-1.1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,216A
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/398,456
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,333
; FILING DATE: 28-AUG-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lippert, Nels T.
; REGISTRATION NUMBER: 25,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-819-8200
; TELEFAX: 212-354-8113
; TELEX: 126201
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2135 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13032
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 514 to 1806
; IDENTIFICATION METHOD: E
US-08-660-216A-3

Query Match 0.5%; Score 19; DB 1; Length 2135;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 AAGTCCAGAAGGAAGCTCG 854
|||||
Db 1504 AAGTCCAGAAGGAAGCTCG 1522

RESULT 8

US-08-211-430-1
; Sequence 1, Application US/08211430
; Patent No. 5763166
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC SEQUENCE OF THE GENE ASSOCIATED WITH

;
; TITLE OF INVENTION: PEPTIDE SEQUENCES, DIAGNOSTIC APPLICATIONS.
; NUMBER OF SEQUENCES: 32
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,430
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6314 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: foetal brain cells
; US-08-211-430-1

Query Match 0.5%; Score 19; DB 1; Length 6314;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3515 CTGTGATCCTGGAGAACT 3533
Db 1223 CTGTGATCCTGGAGAACT 1241

RESULT 9
US-07-988-273-1
; Sequence 1, Application US/07988273
; Patent No. 5512434
; GENERAL INFORMATION:
; APPLICANT: AARONSON, Stuart A.
; APPLICANT: BOTTARO, Donald P.
; APPLICANT: ISHIBASHI, Toshio
; APPLICANT: MIKI, Toru
; TITLE OF INVENTION: EXPRESSION CLONING OF A HUMAN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/988,273
; FILING DATE: 19921214
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/182 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 29..586
; US-07-988-273-1

Query Match 0.5%; Score 18; DB 1; Length 861;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 GGCTGGCGGACTTCATTGA 476
Db 340 GGCTGGCGGACTTCATTGA 357

RESULT 10
PCT-US93-12019-1
; Sequence 1, Application PC/TUS9312019
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: EXPRESSION CLONING OF A HUMAN
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12019
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,273
; FILING DATE: 14-DEC-1992
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 29..586
; PCT-US93-12019-1

Query Match 0.5%; Score 18; DB 5; Length 861;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 GGCTGGCGGACTTCATTGA 476
Db 340 GGCTGGCGGACTTCATTGA 357

RESULT 11
US-09-054-680-3
; Sequence 3, Application US/09054680
; Patent No. 6387694
; GENERAL INFORMATION:
; APPLICANT: McKinney, John D.
; APPLICANT: Jacobs Jr., William R.
; TITLE OF INVENTION: MYCOBACTERIAL ISOCITRATE LYASE GENE AND USES THEREOF
; FILE REFERENCE: 96700/474
; CURRENT APPLICATION NUMBER: US/09/054,680
; CURRENT FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1287
; TYPE: DNA
; ORGANISM: Mycobacterium smegmatis
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1)..(1287)

US-09-054-680-3

Query Match 0.5%; Score 18; DB 4; Length 1287;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 GAAGTTCAGAGGAAC 852
|||||
DB 990 GAAGTTCAGAGGAAC 1007

RESULT 12

US-08-743-637B-20
; Sequence 20, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
US-08-743-637B-20

Query Match 0.5%; Score 18; DB 2; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2396 TTTCATCCCGAGGAA 2413
|||||
DB 104 TTTCATCCCGAGGAA 121

RESULT 13

US-08-526-840B-20

; Sequence 20, Application US/08526840B
; Patent No. 6001564

; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
US-08-526-840B-20

Query Match 0.5%; Score 18; DB 3; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2396 TTTCATCCCGAGGAA 2413
|||||
DB 104 TTTCATCCCGAGGAA 121

RESULT 14
US-08-748-170A-3
; Sequence 3, Application US/08748170A
; Patent No. 6048688
; GENERAL INFORMATION:
; APPLICANT: Korth, Kevin G.
; APPLICANT: Heathcock, Sarah E.
; APPLICANT: Huard, Linda S.
; TITLE OF INVENTION: Method for Detecting Pseudomonas
; TITLE OF INVENTION: aeruginosa Using Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia

COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/748,170A
FILING DATE: 12-NOV-1996

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,697
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Warren, William L.
REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 11302-0350
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2760 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1..2760

OTHER INFORMATION: /note= "Found within the published
OTHER INFORMATION: sequence of the P. aeruginosa Exotoxin A gene; GenBank access
OTHER INFORMATION: number K01397; NCBI sequence ID: 151215"
US-08-748-170A-3

Query Match 0.5%; Score 18; DB 3; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2396 TTTCATCCCCAGCGGAA 2413
Db 104 TTTCATCCCCAGCGGAA 121

RESULT 15

US-09-047-148-1
Sequence 1, Application US/09047148
Patent No. 6086900

GENERAL INFORMATION:
APPLICANT: Draper, Rockford
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR USING
TITLE OF INVENTION: MEMBRANE-PENETRATING PROTEINS TO CARRY MATERIALS ACROSS
TITLE OF INVENTION: CELL MEMBRANES
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/047,148

FILING DATE: Concurrently Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/042,056
FILING DATE: 26-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSF:072
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2760 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-047-148-1

Query Match 0.5%; Score 18; DB 3; Length 2760;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2396 TTTCATCCCCAGCGGAA 2413
Db 104 TTTCATCCCCAGCGGAA 121

Search completed: November 30, 2002, 17:05:28
Job time : 162 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 16:55:19 ; Search time 111 Seconds

(without alignments)
13749.662 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 3963

Sequence: 1 atgcaagacgacagcataga.....tggagagtgtgaactttaa 3963

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 341543 seqs, 192557720 residues

Word size : 0

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	787	19.9	2059	10	US-09-764-853-218
2	348	8.8	702	10	US-09-764-853-397
3	177	4.5	355	10	US-09-563-817-656
4	21	0.5	281	10	US-09-294-093B-4682
5	19	0.5	202	10	US-09-783-590-7307
6	19	0.5	360	10	US-09-745-288-17
7	19	0.5	1228	10	US-09-974-300-244
8	19	0.5	1645	9	US-09-974-298-31
9	19	0.5	1645	10	US-09-919-172-1
10	19	0.5	1823	10	US-09-925-299-231
11	19	0.5	2494	10	US-09-799-983-3
12	19	0.5	2511	10	US-09-799-983-5
13	19	0.5	2682	10	US-09-799-983-1
14	19	0.5	3447	10	US-09-815-242-7797
15	19	0.5	6314	10	US-09-954-456-319
16	19	0.5	6314	10	US-09-954-456-1230
17	19	0.5	9990	10	US-09-767-088A-2
18	19	0.5	9990	10	US-09-767-088A-15
19	19	0.5	10322	9	US-09-764-868-1471

c 20	18	0.5	178	10	US-09-864-761-23722	Sequence 23722, A
c 21	18	0.5	222	10	US-09-757-781-56	Sequence 56, Appl
c 22	18	0.5	230	10	US-09-864-761-22381	Sequence 22381, A
c 23	18	0.5	268	10	US-09-878-574-8528	Sequence 8528, Ap
c 24	18	0.5	277	10	US-09-878-574-5638	Sequence 5638, Ap
c 25	18	0.5	291	10	US-09-764-847-1668	Sequence 1668, Ap
c 26	18	0.5	291	10	US-09-764-847-1669	Sequence 1669, Ap
c 27	18	0.5	411	10	US-09-960-352-4730	Sequence 4730, Ap
c 28	18	0.5	439	10	US-09-815-343-1400	Sequence 1400, Ap
c 29	18	0.5	461	10	US-09-867-701-3100	Sequence 3100, Ap
c 30	18	0.5	476	10	US-09-864-761-5612	Sequence 5612, Ap
c 31	18	0.5	492	10	US-09-867-701-5809	Sequence 5809, Ap
c 32	18	0.5	501	10	US-09-974-300-2396	Sequence 2396, Ap
c 33	18	0.5	567	10	US-09-864-761-15365	Sequence 15365, A
c 34	18	0.5	589	10	US-09-864-761-13612	Sequence 13612, A
c 35	18	0.5	593	10	US-09-864-761-6991	Sequence 6991, Ap
c 36	18	0.5	624	10	US-09-764-847-56	Sequence 56, Appl
c 37	18	0.5	999	10	US-09-887-576-462	Sequence 462, App
c 38	18	0.5	1191	9	US-09-984-245-60	Sequence 60, Appl
c 39	18	0.5	1287	12	US-10-029-715-3	Sequence 3, Appl1
c 40	18	0.5	1333	9	US-09-984-245-110	Sequence 110, App
c 41	18	0.5	1778	12	US-10-052-586-277	Sequence 277, App
c 42	18	0.5	1815	9	US-09-905-291A-136	Sequence 136, App
c 43	18	0.5	1815	10	US-09-909-320-136	Sequence 136, App
c 44	18	0.5	1815	10	US-09-909-088B-136	Sequence 136, App
c 45	18	0.5	1815	12	US-10-052-586-19	Sequence 19, Appl
c 46	18	0.5	1839	10	US-09-758-259-11	Sequence 11, Appl
c 47	18	0.5	1936	10	US-09-764-877-2378	Sequence 2378, Ap
c 48	18	0.5	2005	10	US-09-887-855-1	Sequence 1, Appl1
c 49	18	0.5	2387	10	US-09-815-242-444	Sequence 444, Ap
c 50	18	0.5	2400	9	US-10-108-605-244	Sequence 244, App
c 51	18	0.5	2715	10	US-09-815-242-8309	Sequence 8309, Ap
c 52	18	0.5	2760	10	US-09-452-599-20	Sequence 20, Appl
c 53	18	0.5	2815	10	US-09-312-762A-7	Sequence 7, Appl1
c 54	18	0.5	3451	10	US-09-452-599-18	Sequence 18, Appl
c 55	18	0.5	3723	10	US-09-841-132-484	Sequence 484, App
c 56	18	0.5	3881	9	US-09-954-531-975	Sequence 975, App
c 57	18	0.5	3881	9	US-09-954-531-1374	Sequence 1374, Ap
c 58	18	0.5	3881	10	US-09-967-768A-279	Sequence 279, Appl
c 59	18	0.5	7071	12	US-10-002-600-28	Sequence 28, Appl
c 60	18	0.5	10330	9	US-10-001-189-68	Sequence 68, Appl
c 61	18	0.5	11871	10	US-09-070-927A-171	Sequence 171, App
c 62	17	0.4	142	10	US-09-864-761-21822	Sequence 21822, A
c 63	17	0.4	146	10	US-09-878-574-10053	Sequence 10053, A
c 64	17	0.4	163	10	US-09-864-761-27623	Sequence 27623, A
c 65	17	0.4	167	10	US-09-864-761-26492	Sequence 26492, A
c 66	17	0.4	204	10	US-09-864-761-31004	Sequence 31004, A
c 67	17	0.4	236	10	US-09-923-876-3383	Sequence 3383, Ap
c 68	17	0.4	247	10	US-09-923-876-3035	Sequence 3035, Ap
c 69	17	0.4	270	10	US-09-878-574-8717	Sequence 8717, Ap
c 70	17	0.4	273	10	US-09-923-876-1124	Sequence 1124, Ap
c 71	17	0.4	273	10	US-09-878-574-7760	Sequence 7760, Ap
c 72	17	0.4	273	10	US-09-878-574-8764	Sequence 8764, Ap
c 73	17	0.4	275	10	US-09-878-574-13629	Sequence 13629, A
c 74	17	0.4	291	10	US-09-294-093B-1554	Sequence 1554, Ap
c 75	17	0.4	291	10	US-09-867-701-10142	Sequence 10142, A
c 76	17	0.4	310	10	US-09-864-761-20998	Sequence 20998, A
c 77	17	0.4	315	10	US-09-864-761-28967	Sequence 28967, A
c 78	17	0.4	351	10	US-09-983-965-5200	Sequence 5200, Ap
c 79	17	0.4	363	10	US-09-878-574-2603	Sequence 2603, Ap
c 80	17	0.4	369	10	US-09-864-761-5306	Sequence 5306, Ap
c 81	17	0.4	375	10	US-09-878-574-1729	Sequence 1729, Ap
c 82	17	0.4	377	10	US-09-960-352-14320	Sequence 14320, A
c 83	17	0.4	380	10	US-09-864-761-5522	Sequence 5522, Ap
c 84	17	0.4	393	10	US-09-864-761-4245	Sequence 4245, Ap
c 85	17	0.4	403	10	US-09-880-107-958	Sequence 958, App
c 86	17	0.4	413	10	US-09-960-352-11892	Sequence 11892, A
c 87	17	0.4	440	10	US-09-983-965-3097	Sequence 3097, Ap
c 88	17	0.4	447	10	US-09-983-965-201	Sequence 201, App
c 89	17	0.4	458	10	US-09-864-761-14423	Sequence 14423, A
c 90	17	0.4	470	10	US-09-864-761-5100	Sequence 5100, Ap
c 91	17	0.4	474	10	US-09-864-761-10987	Sequence 10987, A
c 92	17	0.4	485	10	US-09-864-761-14477	Sequence 14477, A

c 969	15	0.4	458	10	US-09-864-761-1187	Sequence 1187, Ap
c 970	15	0.4	458	10	US-09-867-701-729	Sequence 729, App
c 971	15	0.4	459	10	US-09-864-761-3743	Sequence 3743, Ap
c 972	15	0.4	460	10	US-09-864-761-1245	Sequence 1245, Ap
c 973	15	0.4	460	10	US-09-954-456-1096	Sequence 1096, Ap
c 974	15	0.4	462	10	US-09-287-070-3	Sequence 3, Appl1
c 975	15	0.4	463	10	US-09-864-761-6404	Sequence 6404, Ap
c 976	15	0.4	463	10	US-09-864-761-16294	Sequence 16294, A
c 977	15	0.4	464	10	US-09-770-444-353	Sequence 353, App
c 978	15	0.4	464	10	US-09-864-761-14124	Sequence 14124, A
c 979	15	0.4	464	10	US-09-960-352-5924	Sequence 5924, Ap
c 980	15	0.4	465	10	US-09-864-761-1396	Sequence 1396, Ap
c 981	15	0.4	465	10	US-09-864-761-4006	Sequence 4006, Ap
c 982	15	0.4	465	10	US-09-864-761-11249	Sequence 11249, A
c 983	15	0.4	465	10	US-09-864-761-17469	Sequence 17469, A
c 984	15	0.4	466	10	US-09-864-761-5761	Sequence 5761, Ap
c 985	15	0.4	466	10	US-09-783-590-10805	Sequence 10805, A
c 986	15	0.4	467	10	US-09-770-444-274	Sequence 274, App
c 987	15	0.4	467	10	US-09-864-761-16108	Sequence 16108, A
c 988	15	0.4	468	10	US-09-770-444-250	Sequence 250, App
c 989	15	0.4	468	10	US-09-833-381-257	Sequence 257, App
c 990	15	0.4	468	10	US-09-924-035A-210	Sequence 210, App
c 991	15	0.4	470	10	US-09-864-761-5836	Sequence 5836, Ap
c 992	15	0.4	470	10	US-09-864-761-14044	Sequence 14044, A
c 993	15	0.4	470	10	US-09-867-701-1599	Sequence 1599, Ap
c 994	15	0.4	471	10	US-09-815-242-2901	Sequence 2901, Ap
c 995	15	0.4	471	10	US-09-917-800A-910	Sequence 910, App
c 996	15	0.4	471	10	US-09-833-381-1387	Sequence 1387, Ap
c 997	15	0.4	471	10	US-09-867-701-791	Sequence 791, App
c 998	15	0.4	472	10	US-09-867-444-146	Sequence 146, App
c 999	15	0.4	472	10	US-09-770-444-146	Sequence 3901, Ap
c1000	15	0.4	472	10	US-09-864-761-4564	Sequence 4564, Ap

ALIGNMENTS

RESULT 1						
US-09-764-853-218						
Sequence 218, Application US/09764853						
Patent No. US20020090672A1						
GENERAL INFORMATION:						
APPLICANT: Rosen et al.						
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies						
FILE REFERENCE: PJZ06						
CURRENT APPLICATION NUMBER: US/09/764,853						
CURRENT FILING DATE: 2001-01-17						
Prior application data removed - consult PALM or file wrapper						
NUMBER OF SEQ ID NOS: 939						
SOFTWARE: PatentIn Ver. 2.0						
SEQ ID NO 218						
LENGTH: 2059						
TYPE: DNA						
ORGANISM: Homo sapiens						
US-09-764-853-218						
Query Match	19.9%	Score 787;	DB 10;	Length 2059;		
Best Local Similarity	100.0%;	Pred. No. 0;				
Matches 787;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
Qy 3177	AGAAAGACAAAGAGCCCTACAGGACGCTTTTCCGACCAATTTCTCGAGGCTCC	3236				
Db 9	AGAAAGACAAAGAGCCCTACAGGACGCTTTTCCGACCAATTTCTCGAGGCTCC	68				
Qy 3237	TGGGGATATGCTAGCTCATGAGGGGCCCTCTGTCGGCTGAGCTGAAGTGAGTGTTT	3296				
Db 69	TGGGGATATGCTAGCTCATGAGGGGCCCTCTGTCGGCTGAGCTGAAGTGAGTGTTT	128				
Qy 3297	ACCCGCCCGGAGCTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCTCCCA	3356				
Db 129	ACCCGCCCGGAGCTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCTCCCA	188				
Qy 3357	CAAGATGCTGTCAGGAGACCGGAGTCCACTCTCTGCTCAATTGACCCACTCACTACGC	3416				

Db 189	CAAGATGCTGCTGAGGAGACCGAGTCCACTCTCTGCTCATTGACCCACTCACTACGC	248				
Qy 3417	CGAGCGAGGACCTATTAAGTGCATCGTACCAACAACCGGCGAGAAATCTTTTAGTCT	3476				
Db 249	CGAGCGAGGACCTATTAAGTGCATCGTACCAACAACCGGCGAGAAATCTTTTAGTCT	308				
Qy 3477	GGAGCTCTCTGTAGTACCCAAAGAGTGAAGAACCTGTGTATCTCTGGAGAACTACA	3536				
Db 309	GGAGCTCTCTGTAGTACCCAAAGAGTGAAGAACCTGTGTATCTCTGGAGAACTACA	368				
Qy 3537	GAACTGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCC	3596				
Db 369	GAACTGCGGTGTTCCCGAAGGCCACCCCGTGAGACTGGAGTGGCGGTGATAGGCATGCC	428				
Qy 3597	CCACCTGCTGTTCTACTGGAAGAACATGAGACCATCCCTTGCACGAGAGAGGAT	3656				
Db 429	CCACCTGCTGTTCTACTGGAAGAACATGAGACCATCCCTTGCACGAGAGAGGAT	488				
Qy 3657	CAGTATGCACGAGACACACAGGGTATGCTGCTTCTCATTGAGCCAGCCAGAAATC	3716				
Db 489	CAGTATGCACGAGACACACAGGGTATGCTGCTTCTCATTGAGCCAGCCAGAAATC	548				
Qy 3717	AGAGCGTGATGTTACAGTTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGACTGC	3776				
Db 549	AGAGCGTGATGTTACAGTTGTCAGCCAAAGATGAAGCCGGCATCGTGTGCGACTGC	608				
Qy 3777	CAGGCTGGATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGCTGTCCGGCC	3836				
Db 609	CAGGCTGGATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGCTGTCCGGCC	668				
Qy 3837	CAGTGGCAGTCGGTACGGATCTCTCACCAGTAAAGGACTTTGACATATTTTCGCTTTTC	3896				
Db 669	CAGTGGCAGTCGGTACGGATCTCTCACCAGTAAAGGACTTTGACATATTTTCGCTTTTC	728				
Qy 3897	CTCATGGAAGACGATGTTATTCATGCTCTTCTCGGAGTGTAGTGAGAGTGATGA	3956				
Db 729	CTCATGGAAGACGATGTTATTCATGCTCTTCTCGGAGTGTAGTGAGAGTGATGA	788				
Qy 3957	ACTTTAA 3963					
Db 789	ACTTTAA 795					

RESULT 2
US-09-764-853-397
Sequence 397, Application US/09764853
Patent No. US20020090672A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PJZ06
CURRENT APPLICATION NUMBER: US/09/764,853
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 939
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 397
LENGTH: 702
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (587)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: SITE
LOCATION: (661)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: SITE
LOCATION: (670)
OTHER INFORMATION: n equals a.t.g, or c
US-09-764-853-397

```
; OTHER INFORMATION: n = A,T,C or G
US-09-563-817-656

Query Match      4.5%; Score 177; DB 10; Length 355;
Best Local Similarity 99.5%; Pred. No. 1.9e-82;
Matches 227; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  903 GTGCTACTGTGAAGCAAGGAGCTTGAATAATCCCGACAGATATTACATCATGTCAGGCAGG 962
    |||||
DB   1 GTGCTACTGTGAAGCAAGGAGCTTGAATAATCCCGACAGATATTACATCATGTCAGGCAGG 60

QY  963 AAATCTGCATCTACTGACCATTGGGAAAGCCTTTGAAGAGGACACAGGACGCTATTCTCTG 1022
    |||||
DB   61 AAATCTGCATCTACTGACCATTGGGAAAGCCTTTGAAGAGGACACAGGACGCTATTCTCTG 120

QY  1023 CTTTGCTTCTTAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATAGAAGGGGT 1082
    |||||
DB  121 CTTTGCTTCTTAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATAGAAGGGGT 180

QY  1083 TTTCTTCTTCTGACTCAGAAGCGCACCTTAACAAGGAGAGATGAATCG 1130
    |||||
DB  181 TTTCTTCTTCTGACTCAGAAGCGCACCTTAACAAGGAGAGATGAATCG 228

RESULT 4
US-09-294-093B-4682
; Sequence 4682, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4682
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700354793H1
; NAME/KEY: unsure
; LOCATION: 127
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-4682

Query Match      0.5%; Score 21; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  157 GGAGCGGAGGCGCAAGATGAC 177
    |||||
DB  175 GGAGCGGAGGCGCAAGATGAC 195

RESULT 5
US-09-783-590-7307/c
; Sequence 7307, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products
; FILE REFERENCE: PO-16,2C1
```

; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7307
; LENGTH: 202
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (58)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (107)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (152)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (175)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-7307

Query Match 0.5%; Score 19; DB 10; Length 202;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2225 CTCGGGTGTCACCTTGAG 2243
|||||
Db 129 CTCGGGTGTCACCTTGAG 111

RESULT 6
US-09-745-288-17
; Sequence 17, Application US/09745288
; Patent No. US20010018058A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, DavIn C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.44601
; CURRENT APPLICATION NUMBER: US/09/745,288
; CURRENT FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(360)
; OTHER INFORMATION: n = A,T,C or G
US-09-745-288-17

Query Match 0.5%; Score 19; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2275 AGCAAGAAAGCCTCTTAG 2293
|||||
Db 35 AGCAAGAAAGCCTCTTAG 53

RESULT 7
US-09-974-300-244
; Sequence 244, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-244

Query Match 0.5%; Score 19; DB 10; Length 1228;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 984 TCGGGAAGCCTTTGAAGAG 1002
|||||
Db 219 TCGGGAAGCCTTTGAAGAG 237

RESULT 8
US-09-974-298-31
; Sequence 31, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1382961.3
US-09-974-298-31

Query Match 0.5%; Score 19; DB 9; Length 1645;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1627 TCTCTTCACTCAGCCTACT 1645
|||||
Db 25 TCTCTTCACTCAGCCTACT 43

RESULT 9
US-09-919-172-1
; Sequence 1, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS

FILE REFERENCE: PA-0036 US
CURRENT APPLICATION NUMBER: US/09/919,172
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/222,469
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 1645
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020119463A1 1382961.3
US-09-919-172-1

Query Match 0.5%; Score 19; DB 10; Length 1645;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1627 TCTCTTCACTCAGCCAACT 1645
|||||
Db 25 TCTCTTCACTCAGCCAACT 43

RESULT 10
US-09-925-299-231
Sequence 231, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 231
LENGTH: 1823
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1593)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1714)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-231

Query Match 0.5%; Score 19; DB 10; Length 1823;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3737 TGTGAGCCCAAGAAATGAAGC 3755
|||||
Db 1408 TGTGAGCCCAAGAAATGAAGC 1426

RESULT 11
US-09-799-983-3
Sequence 3, Application US/09799983
Patent No. US20010029033A1
GENERAL INFORMATION:
APPLICANT: Shaml, Paul
APPLICANT: Parker, Charles

TITLE OF INVENTION: NOVEL GENE RNO UPREGULATED IN LEUKEMIA CELLS BY NITRIC OXIDE
FILE REFERENCE: 1321.2.51
CURRENT APPLICATION NUMBER: US/09/799,983
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: US 60/186,971
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 2494
TYPE: DNA
ORGANISM: Homo sapiens
US-09-799-983-3

Query Match 0.5%; Score 19; DB 10; Length 2494;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2797 CAACATGATGAGATCCCA 2815
|||||
Db 442 CAACATGATGAGATCCCA 460

RESULT 12
US-09-799-983-5
Sequence 5, Application US/09799983
Patent No. US20010029033A1
GENERAL INFORMATION:
APPLICANT: Shaml, Paul
APPLICANT: Parker, Charles
TITLE OF INVENTION: NOVEL GENE RNO UPREGULATED IN LEUKEMIA CELLS BY NITRIC OXIDE
FILE REFERENCE: 1321.2.51
CURRENT APPLICATION NUMBER: US/09/799,983
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: US 60/186,971
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 2511
TYPE: DNA
ORGANISM: Homo sapiens
US-09-799-983-5

Query Match 0.5%; Score 19; DB 10; Length 2511;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2797 CAACATGATGAGATCCCA 2815
|||||
Db 459 CAACATGATGAGATCCCA 477

RESULT 13
US-09-799-983-1
Sequence 1, Application US/09799983
Patent No. US20010029033A1
GENERAL INFORMATION:
APPLICANT: Shaml, Paul
APPLICANT: Parker, Charles
TITLE OF INVENTION: NOVEL GENE RNO UPREGULATED IN LEUKEMIA CELLS BY NITRIC OXIDE
FILE REFERENCE: 1321.2.51
CURRENT APPLICATION NUMBER: US/09/799,983
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: US 60/186,971
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2682
TYPE: DNA
ORGANISM: Homo sapiens
US-09-799-983-1

```
Query Match      0.5%; Score 19; DB 10; Length 2682;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2797 CAACATGATGAGATCCCA 2815
   |||||
Db 459 CAACATGATGAGATCCCA 477

RESULT 14
US-09-815-242-7797
; Sequence 7797, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7797
; LENGTH: 3447
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(3447)
US-09-815-242-7797

Query Match      0.5%; Score 19; DB 10; Length 3447;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 CTGGGGCAACCTCCCGGT 811
   |||||
Db 51 CTGGGGCAACCTCCCGGT 69

RESULT 15
US-09-954-456-319
; Sequence 319, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
```

```
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 319
; LENGTH: 6314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-319

Query Match      0.5%; Score 19; DB 10; Length 6314;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3515 CTGTGATCCTGGAGAAACT 3533
   |||||
Db 1223 CTGTGATCCTGGAGAAACT 1241

Search completed: November 30, 2002, 18:41:22
Job time : 185 secs
```

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 30, 2002, 19:07:45 ; Search time 5366 Seconds
(without alignments)
7159.095 Million cell updates/sec

Title: US-09-818-990B-2
Perfect score: 6890
Sequence: 1 MQDDSEASTISQILLRESY.....MESTMYSCSRSVVESDEL 1320

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/uspto.spool/us09818990/runat_26112002_093404_23219/app_query.fasta_1.1479
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1_1.3586 @runat_26112002_093404_23219 -NGPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -THREDOVER=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

em.vi.*
em.htg.hum.*
em.htg.inv.*
em.htg.other.*
em.htg.mus.*
em.htg.pln.*
em.htg.rod.*
em.htg.mam.*
em.htg.vrt.*
em.sy.*
em.htgo.hum.*
em.htgo.mus.*
em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6869	99.7	5707	9	AF328296 Homo sapi
2	6863	99.6	5804	9	AL834247 Homo sapi
3	6842	99.3	5486	9	AL832002 Homo sapi
4	5322	77.2	4683	9	AL833379 Homo sapi
5	2560	37.2	2127	9	AK027343 Homo sapi
6	2299	33.4	5773	9	AF464873 Homo sapi
7	2081.5	30.2	2920	9	AK095512 Homo sapi
8	2042.5	29.6	3147	10	BC027364 Mus muscu
9	1953.5	28.4	136493	2	AC125500 Takifugu
10	1666	24.2	4347	9	AB023209 Homo sapi
11	1553	22.5	160658	9	AC024258 Homo sapi
12	1553	22.5	175265	2	AC063964 Homo sapi
13	1553	22.5	180707	2	AC024460 Homo sapi
14	1476	21.4	1664	9	BC013867 Homo sapi
15	1460	21.2	1580	10	AF205078 Mus muscu
16	1425.5	20.7	1409	9	AK025843 Homo sapi
17	1408.5	20.4	1401	9	AK000845 Homo sapi
18	1379.5	20.0	3757	9	AF151909 Homo sapi
19	1124	16.3	11170	9	AL512429 Human DNA
20	1124	16.3	290625	2	AL442068 Homo sapi
21	1071.5	15.6	151002	2	AC114220 Rattus no
22	874.5	12.7	12699	5	D83390 Gallus gall
23	826	12.0	6492	5	AF159173 Gallus ga
24	810.5	11.8	81940	9	X90568 H.sapiens m
25	807.5	11.7	6000	4	OCU18102 Oryctolagus
26	760	11.0	2244	9	AF144477 Homo sapi
27	760	11.0	2273	9	BC005376 Homo sapi
28	760	11.0	2275	9	AF133820 Homo sapi
29	740	10.7	175265	2	AC063964 Homo sapi
30	740	10.7	290625	2	AL442068 Homo sapi
31	704	10.2	2222	10	BC016214 Mus muscu
32	704	10.2	2223	10	AF230979 Mus muscu
33	699	10.1	2320	9	AK096458 Homo sapi
34	682	9.9	1427	10	AF205079 Mus muscu
35	609	8.8	16076	3	AB055862 Procamb
36	609	8.8	53424	3	AB055861 Procamb
37	583	8.5	15244	3	AB026845 Drosophil
38	582	8.4	3028	3	DMTITIN1 AF045775 Drosophil
39	580	8.4	90803	5	AL732421 Zebrafish
40	579	8.4	29280	3	AB079865 Bombyx mo
41	579	8.4	160000	3	AB090307 Bombyx mo
42	577	8.4	83495	3	AC087076 Caenorhab
43	566	8.2	53848	3	AF003131 Caenorhab
44	566	8.2	96468	2	AC006902 Caenorhab
45	566	8.2	299719	2	AC006858 Caenorhab

ALIGNMENTS

RESULT 1

AF328296 5707 bp mRNA linear PRI 06-MAY-2001
LOCUS Homo sapiens myopalladin mRNA, complete cds.
DEFINITION AF328296
ACCESSION AF328296.1 GI:13957726
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5707)
AUTHORS Bang,M.L., Mudry,R.E., McElhinny,A.S., Trombitas,K., Geach,A.J.,
Yamasaki,R., Sorimachi,H., Granzler,H., Gregorio,C.C. and Labelit,S.
TITLE Myopalladin, a novel 145-kilodalton sarcomeric protein with
multiple roles in z-disc and I-band protein assemblies
J. Cell Biol. 153 (2), 413-427 (2001)
JOURNAL
MEDLINE 21206024
PUBMED 11309420
REFERENCE 2 (bases 1 to 5707)
AUTHORS Bang,M.-L.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2000) Structural and Computational Biology,
European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1,
Heidelberg 69117, Germany
LOCATION/Qualifiers
source 1. .5707
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q21.1"
/tissue_type="striated muscle"
488..4450
/codon_start=1
/product="myopalladin"
/protein_id="AAK50625.1"
/db_xref="GI:13957727"
/translation="MQDSEASTSISQLRESYLAEIRHRGNERSRAPSSNPCHF
GSPGAAAGGGDDLPDLFAFLSELDSEVNLAIRINIDPLEKADTEQARKRLSP
DOMKSNLSPFNCDNPRSTSSKEPQAKRPOYCESTQSKVFLNKAADFTBE
LSLFLKSHSKRIKPRACKNHKSLKSONKMOENSSFDLSERRRSPVPIPAD
TRONEVNHAEQOEAKREAOAASEAGDTPGSSPSLSLYEEELGQPPRTOKLR
SREVPETRVQLDCIVGIPPOVRWYCEGKELENSDDIHVQAGNLHSLTIAEBEE
DTGRYSQAHLVGTDSAEIYIEGVSSDSEDPNKEEMRIQKPNVSSPPTTSA
VIPPAPQAGHLVAPQVATIQCCSPTNYQLQDGKPIIAAPVFTKMLQNLASQEG
LVVFECRGAPSKVYEWYREGTLIEDPDFRILQKPRSMABPEICTLVIAEVFAE
DSGCTAGNKGVTSSIAQLHVRGNEDLSNGLHSANSTNLAIEEPQPPHSE
PPSVQPPKPKLEGLVNHNEPRSSRIGLRVHNLPEDDKGSSEASEAGVITROT
PDSEFPNGQAKTPEPSPVKEPPVLAAPKLDSTQLQHLNVLLEQHLQNLPPP
SSPEFPFNMTVNSNAPPATVTSNKQVAPSSQTSFLARPKYFFFTNTATVPAS
SSPVFTLSSTPQIQTIVSKESLLVSHPSVQTSKPGSLSIONEPLPPGTEPPPT
FSIPRGNOFCVSPVPTSPTRIQNPVAFLSVLSPALPIPTNMGIPRSPASMP
SGLAKNTKSPQVNDNTRETKNAVIRDLGKTIKTFSDVRPNQOEYKISFQRLWN
EIEFRLETPVDESDDIOHDEIPTGKCIAPFDKRLKHFRTVEGSPVETCKIVGP
VPKYVFKDKQISKREHCKMRREGDGTCSLHISTSDDDGNITYTMAANPOGRISC
SGHLVQSLPRLRLTSAGHSRQVQRDEKPELQERFREFPHFLOAQDMDVAHER
LCRDCKVSLPPPELWLLNGQVPLPVDASHKMLVRETGVHSLDPLTDQMDAGTYK
IATNKTCQNSFELSLSVAKVKKVLEQLKQNGVPEGHVRLERCVIGMPPPVFY
WKQNETIPTCRISIMHODITGYACLLIQAPAKKSDAGWVTLNLAKEAGIVSCSTARID
IYAQWHHQIIPPMVSRPSGSGSLTSKGLDIFSAFSSMESTWYVSSCSRVVSDBL
1277..1531
/note="Region: Ig-I immunoglobulin domain"
misc_feature 1790..2044
/note="Region: Ig-I immunoglobulin domain"
misc_feature 3299..3565
/note="Region: Ig-I immunoglobulin domain"
misc_feature 3722..3946
/note="Region: Ig-I immunoglobulin domain"
misc_feature 4001..4246
/note="Region: Ig-I immunoglobulin domain"
BASE COUNT 1623 a 1437 c 1294 g 1353 t
ORIGIN

Alignment Scores:
Pred. No.: 2.06e-252 Length: 5707
Score: 6869.00 Matches: 1314
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 6
Query Match: 99.70% Indels: 0
DB: 9 Gaps: 0
US-09-818-990b-2 (1-1320) x AF328296 (1-5707)
Qy 1 MetGlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 488 ATGCAAGACGACACATAGAACTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 547
Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 548 TTAGCTGNAACACACATCGGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCTCCAC 607
Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlyGlnAspLeu 60
Db 608 CTTTGCCATTTCGCGAGCTCTTCTGGGCGCTGAAAGGAGCGGAGGCAAGATGACCTT 667
Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 668 CCAGATCTTTTCAGCTTTTCTGAGCCAAAGAAATTAGACGAAGTGTCAATTTGGCAAGA 727
Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Db 728 CTGGCCATCAATTAGACCCCTTTGGAGNAGCAGATGAACTCAAGCTAGAAAACGACTT 787
Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Db 788 TCTCCTGATCAGATGAAACACTCACCCTAAATTAAGTTTGTAGCCTTAATCTTCCAGGAT 847
Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Db 848 AACCTCGAAGTCCACACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAGGCCACAGTAT 907
Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Db 908 TGTTCGNAACCCAGTCCAAAAAGTATTTTAAATTAAGGCTGCCGACTTCATTGAAGAG 967
Qy 161 LeuSerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsn 180
Db 968 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAAGGATTAGACCTCGCTGCTGCAAAAA 1027
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSerPheSer 200
Db 1028 CACAAGAGTAAACTGGGAATCTCAAAACAAAGTTTATGCAGGAAACAGCTCCAGTTTCTCA 1087
Qy 201 AspLeuSerGluArgArgGluArgSerValProIleProIleProAlaAspThrArg 220
Db 1088 GATCTGTGAGAAACAGCAGAAAGATCTTCTGTTCCCATCCCTATCCCTGCGGATACCA 1147
Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
Db 1148 GATAATGAAGTGAATCACGCCCTCGAACAGCAGGAAAGCAAGAGCGGTGAAGCGGAGCAG 1207
Qy 241 AlaAlaSerGluAlaAlaGluGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
Db 1208 GCTGCCAGTAGAGCGGCTGGTGGAGACACTACACACAGGGTCTTCCCTTCATCTCTGTAC 1267
Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
Db 1268 TATGAGAACCTCTGGGGCAACCTCCCGGTTCTACTCAAAAGTTACGGAGCAGAGAGTT 1327
Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
Db 1328 CCAGAAGGAACCTCGAGTACAGTTGATTTGCATAGTGGTAGGAATTTCCACCACCTCAAGTA 1387
Qy 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
Db 1388 AGTGGTACTGTGAAGCCAGGAGCTTTGAAATTTCCCCAGATATTCACATCTGTCAGGCA 1447

Qy	581	ProProSerSerProLysGluPheProPhe***MetThrValLeuAsnSerAsnAlaPro	700
Db	2528	CCTCCTTCATCTCCTAAGGAGTTTCCTTTCAACATGACTGTTTGAACATCCAAATGCTCCC	2587
Qy	701	ProAlaValThrThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeu	720
Db	2588	CCAGCGTGACACATCCCAATGAAGCAGGTGAAGGCTCTTCATCACAGCGTTCAAGCTTG	2647
Qy	721	AlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSer	740
Db	2648	GCCGGCCGAAGTATTCCTCCCTCCACGAACACCCAGCAGCAACTGTGSCCCTTCC	2707
Qy	741	SerSerProValPheThrLeuSerSerThrProGlnThrThrAlaAlaThrValAlaProSer	760
Db	2708	AGCTCTCCGGTGTTCACTTTGAGCAGCACTCTCTAAACTATTTCAGAGCAGCAGTGAAGAAA	2767
Qy	761	GluSerLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIle	780
Db	2768	GAAAGCCCTCTAGTGCTCACCCCTCTGTGCAACCAAACTCTCCAGGAGGCTTTCCATC	2827
Qy	781	GlnAsnGluProLeuProGlyProThrGluProThrProProProPheThrPheSer	800
Db	2828	CAAAATGAGCCACTCCACCAGGCCCAACAGAACCAACACACACACCACTTACACTTTC	2887
Qy	801	IleProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSerProThr	820
Db	2888	ATCCCCAGGGGAACACAGTTTCAGCCCGCTGTGTGTCCCAATTCCTCTCTCTTACC	2947
Qy	821	SerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIle	840
Db	2948	AGCCGGAATCAGAACCCAGTGGCTTTCCTCAGCTCTGTCTTGCCTTCTCTCCCTGCCATC	3007
Qy	841	ProProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeu	860
Db	3008	CCACCCACAATGCCATGGGGTGCCTAGAGTGCACCATCCATGCCATCCACGGATTAC	3067
Qy	861	AlaLysLysAsnThrLysSerProGlnProValAsnAspAsnIleArgGluThrLys	880
Db	3068	CGGAAGAAAAATCAAAAGTCTCTCAACAGTGAATGATGATAACATTCGTGAAACTAAG	3127
Qy	881	AsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAsnGln	900
Db	3128	AACCGCATGATTCGAGACTTGGGAAAAAATAACTTTTCAGTGATGTCAGACCAACACAG	3187
Qy	901	GlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeu	920
Db	3188	CAGGAGTACAAAATTCAAAGCTTTGAGCAGAGCGTGAATGAATAGAGTTTCGCTTG	3247
Qy	921	GluArgThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGly	940
Db	3248	GAACGTACTCTGTGTATGAATCAGATGATGAATTAACATCATGATGATGATCCCCACGGC	3307
Qy	941	LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluClySer	960
Db	3308	AAGTGTATTGCTCCCATCTTTCAGACAGACATCAAGCACTTCCGGGTCAACAAGAGCTCT	3367
Qy	961	ProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLys	980
Db	3368	CCATGCATTCACCTGCAAAATTTGGGTACCTGTGTCCAAAAGTTTACTGGTTCAA	3427
Qy	981	AspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluClyAspGly	1000
Db	3428	GATGGGAAGCAGATTTCTAAGACAAATGAGCACTGCAAAATCAGCCGAGAGGAGATGGG	3487
Qy	1001	ThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMet	1020
Db	3488	ACATGCTCTCTGCACATTAATCCACTACCACTGATGACGATGGCACTTACACCATCATG	3547
Qy	1021	AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro	1040
Db	3548	GCACCAACCCCGAGGGGAATCAGCTGTCTTGGCCACTTATGGTCAAAAGTTTGCC	3607
Qy	1041	IleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGlu	1060

```
|||||
Db 3608 ATTCCGAGTCGCTAACTCTGCTGGTCAGTCTCACAGGGAAGATCCGAGTCAAGAA 3667
Qy 1061 ArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGly 1080
Db 3668 AGACACAAGAGCCCCACAGGAACGCTTTTTCGAGCACACATTTTCCTGAGGCTCCCTGGG 3727
Qy 1081 AspMetValAlaHisGluGluArgLeuGluCysArgLeuAspCysLysValSerGlyLeuPro 1100
Db 3728 GATATGTAGTCTCATGAGGGCGCTCTGTGGCTGACATGTAGGTGAGTGGTTTACCG 3787
Qy 1101 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLys 1120
Db 3788 CCCCGGAGCTGACATGGCTACTCAATGGCAACCTGTGCTACACAGATGCTCCGCCACAAG 3847
Qy 1121 MetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArgAsp 1140
Db 3848 ATGCTGTGTCAGGAGAGCGGAGTCCTCTGTCTCATTTGACCCACTCACTCAGCGCGAC 3907
Qy 1141 AlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGlu 1160
Db 3908 GCAGGACCTAATAGTCATGCTGCTACCAACAAACCGGCGAGAAATCTTTTAGTCTGGAG 3967
Qy 1161 LeuSerValValAlaLysGluValLysLysAlaProValLleLeuGluLysLeuGlnAsn 1180
Db 3968 CTCCTCTAGTACCAAGAGGTCAAGAAAGCACCTGTGATCTCTGGAGAACTACAGAAC 4027
Qy 1181 CysGlyValProGluGluHisProValArgLeuGluCysArgValLleGlyMetProPro 1200
Db 4028 TGGCGGTGTTCCGGAAGCCACCCCGTGAAGTGGAGTGGCGGTGATGAGCATGCCGCCA 4087
Qy 1201 ProValPheTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSer 1220
Db 4088 CCTGTGTTCTACTCGAAGAAGACAAATGAGACCATCCCTGTCACACAGAGAGATCAGT 4147
Qy 1221 MethisGlnAspThrThrGlyTyrAlaCysLeuLeuLeuGlnProAlaLysLysSerAsp 1240
Db 4148 ATGCACCAAGGACACACAGGGTATGCTGCTCTCTCATTCAGCCAGCCCAAGAAATCAGAC 4207
Qy 1241 AlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArg 1260
Db 4208 GCTGGATGGTACAGCTGTGTCAGCCAAAGTGAAGCCGGCATCGTGTGTCGACATGCCAGG 4267
Qy 1261 LeuAspIleTrpAlaGlnTrpHisGlnIleProProMetSerValArgProSer 1280
Db 4268 CTGGATATATACGCTCAGTGGCACCATCAGATCCACCGCCCATGCTGTGCGGCCCATG 4327
Qy 1281 GlySerArgTrpGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerSer 1300
Db 4328 GGCAGTCGTCACGATCTCTCACCAGTAAAGGACTTGACATATTTTCTGCTTTTCTCCTCC 4387
Qy 1301 MetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSerAspGluLeu 1320
Db 4388 ATGGAAAACGAGTGGTGTATTTCATGCTCTCTCGGAGTGTAGTGGAGATGATGAACCT 4447

RESULT 2
HSM805275 5804 bp mRNA linear PRI 12-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp451E012 (from clone DKFZp451E012).
DEFINITION AL834247
ACCESSION AL834247.1 GI:21739789
VERSION
KEYWORDS
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5804)
AUTHORS Wambutt R., Heubner D., Mewes H.W., Weil B. and Wiemann S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
```

consortium of the German Genome Project.
This clone (DKFZp451E012) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
Location/Qualifiers
source
1..5804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="DKFZp451E012"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..5804
/gene="DKFZp451E012"
CDS
1..4227
/gene="DKFZp451E012"
/note="similarity to actin-associated protein palladin
(Mus musculus)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD38923.1"
/db_xref="GI:21739790"
/translation="YCEKTEGGADKCSGGDRLLSSFLQPSLLKRLTPHSLMTAT
LYFGREFFSPSPVLSKSPVITITAVELCDWISSGSTKKESMODDSEASTSISQ
LURESYLAETRHGNNERAEPSNPHFGSPGAEGGGDDLDLUSAFUSBEL
DESVNLARLAINYDPLEKADETOARKRLSPDQMKHSPNLSFEPNQQDNPRTSSKE
SPOEAKRPQVCSCTSKVKFLNKAADIFBELSLFKSHSKRIIPRACKNHKSKLSQ
NKVMQNSSSSDLSRRERSVPIPIPADTRDNEVHALAQEAKRRARAEQASAA
QGDTPCSPSSSLYEEPLGQPPFTOKLRSREVPEGTVOLDICVVGIPPOVRKYC
EGKELNSPDHIVQAGNLHSLTIAEPEDTGRYSCFASNIYGTSTSAEYIEGVS
SSDEGDPNKEENRIQKNESVPTTSAVIPAPVQAHLVAQPRVATIQOQSPT
NYLGLDGKPIITAAQFTKLNLSAQGLVVEFCRVKGAQSPKVEWREGLIEDS
PDFRILOKPPRMAPEEICTLVIAEFADSGCTCTASNKYGTVSSIAQLHVRNE
DLSNGSLNSATNTLAAETPOQSPHPSEVPQPPKPLEGLVLNNEPRSSRI
GLRVHNLPEDDKSGSEASGAVVTTROTTPDSFORFENGQATKTPEPSVPKEPPV
LAKPKLDSTQLQQLHQLVLEQLQNPSPSPKPEFESWTVLNSNAPVATISSQV
KAPSTQTFSLARPIFFPTNTTAAATVAPSSPVFTLSPTQITQTVSKESLULSHP
SVQTSKPGGLSTQNEPLPFPTEPTTPTTIPSGNQFPQPCRPVPSPTRIQNP
VPLKSLTSLPALPAINAMLPSPASMPSPQSLAKNTKSPQVADDNDIODEPTCKC
RDLGKXITFSDVRPNQOEVKISSFEQRLMNEIFRLERTPVDESDEIOHDEPTCKC
IAPIDFKLKHFEVTPGSPVETCKTIGVIPVKVWFQKQKQISKNRCKMRREGDG
TCSLHLESTSDDDGNTTMAANPQGRISCSGHLVQSLPISRLSLTSAGSHRGSRV
QERDKPLQERFRFFLQAGDMVAHRELCRLDCKVSGLPPELTLWLLGQPVLPD
APHKMLVRETGVHSLLDPLTORDACTYKCIATNKTGQNSFSELSVAVAKKAPVI
LEKLQCGVPEGHPVRLCVRIGMPVPVFWKDNETIPTCTRERISNHQDTTGYACIL
IOPAKKSDAGWYTLAKNEAGIVSCATRLDIYAQWHHIQIPPPMSVRPSGRYGLTSK
GLDIFSASFSSMESTWYVYSCSSRSRVESDEL"
polyA_signal 5766..5771
/gene="DKFZp451E012"
polyA_site 5783
/gene="DKFZp451E012"
BASE COUNT 1650 a 1460 c 1322 g 1372 t
ORIGIN

Alignment Scores:
Pred. No.: 3..55e-252 Length: 5804
Score: 6863.00 Matches: 1313
Percent Similarity: 99.47% Conservative: 0
Best Local Similarity: 99.47% Mismatches: 7
Query Match: 99.61% Indels: 0
DB: 9 Gaps: 0

US-09-818-990b-2 (1-1320) x HSM805275 (1-5804)

Qy 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuArgGluSerTyr 20
|||||
Db 265 ATGCAAGACACACATAGAGGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGATCAT 324
Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40

325	Db	TTAGCTGAAACCAAGACATCGGGAAACAATGAGAGAGTGTGAGCGGAGCCCTCTCTCCAAC	384
41	Qy	ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeu	60
385	Db	CCTTGCCATTTCGGCAGTCTCTCTGGGCGCTGAAGGAGCGGAGCCAGATGACCTT	444
61	Qy	ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg	80
445	Db	CCAGATCTTTCAGCCCTTCTGAGCCAAGAAGAAATTTAGACGAAAGTGTCAATTTTGGCAAGA	504
81	Qy	LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaAtrLysArgLeu	100
505	Db	CTGCCCATCAATTACGACCCCTTTGGAGAAGGCAGATGAACTCAAGCTAGAAAAACGACTT	564
101	Qy	SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp	120
565	Db	TCTCTCATCATGATGAACACACTCACCTAATTTAAGTTTGGCCTAACTTCTGCCAGAT	624
121	Qy	AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr	140
625	Db	AACCTCTGAAGTCCCAACAGCTCTAAGAAAGAGCCCCAGGAGGCAAAAAGGCCACAGTAT	684
141	Qy	CysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu	160
685	Db	TGTTTCTGAACCCAGTCCAAAAAGTATTTTAAATAAAGTGCGCAGCTTCATTGAAGAG	744
161	Qy	LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn	180
745	Db	CTATCTCCCTTTTCAATCCCACTCCAAAGAGATTAGACCTGTGCTCGCAAAAC	804
181	Qy	HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer	200
805	Db	CACAAGAGTAACTGGAACTCTCAAAACAAGTTATGCAGGAAACAGCTCCAGTTCTCA	864
201	Qy	AspLeuSerGluArgArgGluArgSerValProIleProIleProAlaAspThrArg	220
865	Db	GATCTGTCAAGAACGACGAGAAAGATCTTCTGTTCCTCATCCCTATCCCTCGCGGATACCCAGG	924
221	Qy	AspAsnGluValAsnHisAlaLeuGluGlnGlnAlaLysArgArgGluAlaGluGln	240
925	Db	GATTAATGAATGAATCAGCCCTGTGAACACAGCAAGAACCAAGAGCGGTGAAGCGGAGCAG	984
241	Qy	AlaAlaSerGluAlaLagLysAspThrThrProGlySerSerProSerSerLeuTyr	260
985	Db	GCTGCCAGTGAGCGGCTGGTGGAGACACTACACAGGGTCTTCCCTTCATCTCTGTAC	1044
261	Qy	TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal	280
1045	Db	TATGAAGAACCTCTGGGGCAACCTCCCGGTTCACCTCAAAAGTTACGAGCAGAGAAGTT	1104
281	Qy	ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGluVal	300
1105	Db	CCAGAGGAACCTCGATPACAGTTGGAAATGCATAGTGTAGGAATTCACCACCTCAAGTA	1164
301	Qy	ArgTrpTyrCysGluLysGluLeuGluAsnSerProAspIleHisIleValGlnAla	320
1165	Db	AGGTGGTACTCTGAGGCAAGGAGCTTGAATTTCCCCAGATATTCACATCGTCCAGCA	1224
321	Qy	GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer	340
1225	Db	GGAATCTGCACCTCACTGCACCATTCGCGAAGCGCTTGAAGAGGACACAGGACGCTATTTC	1284
341	Qy	CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly	360
1285	Db	TGCTTTTCTCTAACATCTATGGACAGATTCGACTCTGCTGAGATTTATATAGAAGG	1344
361	Qy	ValSerSerSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys	380
1345	Db	GTTTCTTCTGACTCAGAAAGCGCACCTTACAGGAGAGAGATGCATTCGAATCCAGAG	1404
381	Qy	ProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValProGln	400

Db	1405	CCAAATGAGGTGCATCTCTCTCCCACTACTCTG	CAGTCACTTCCTCCAGCAGTACCCGAA	1465
Qy	401	AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr		420
Db	1465	GCCCAGCATTTGGTGGCCCAACCTCGTGTGGCAACCATCCAGCAGTGTCAAAGCCCCACC		1524
Qy	421	AsnTyrIleuGlnGlyLeuaspGlyLysProIleIleAlaAlaProValPheThrLysMet		440
Db	1525	AATTACTTGCAGGGATTGGATGGAAACACTATCATTTGCAGCTCTCTGTGTTTACAAGATG		1584
Qy	441	LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly		460
Db	1585	CTACAAATTTTGCAGCTCTGAGGGTCAGCTGGTTGCTCTTTGAATGCAGAGTAAAGGA		1644
Qy	461	AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluaspSerProAsp		480
Db	1645	GCTCCATCTCCTAAGGTTGAGTGTATAGAGAAGGGACTTTTAATAGAAGATCTCCAGAT		1704
Qy	481	PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu		500
Db	1705	TTTATAGGATTTTACAGAAAAACCTCCATCCATGGCAGAGCCAGAGAGATTTGCACCTTG		1764
Qy	501	ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys		520
Db	1765	GTCATTGCTCAGGTGTTGCGAAGAGATTCTGGTGCTTCCATGTACTGACGAAGCAACAAA		1824
Qy	521	TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluaspLeuSerAsn		540
Db	1825	TACGGCACAGTGTCAAGCATTCGACAGCTGCACGTGAGAGGAATATGAGGACTCTCAGC		1884
Qy	541	AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnPro		560
Db	1885	AACGGGTCTCTTCACTCAGCCAACTCTACCACCACTGGCAGCTATTGAGCCACAGGCC		1944
Qy	561	SerProProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGly		580
Db	1945	TCCCCACCCCACTCAGAGCCTCCATCTGTGGAAACACCCCCCAAACTCGAGGGG		2004
Qy	581	ValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArgValHisPhe		600
Db	2005	GTTCTGGTGAACCAATGAGCCCCCGTCCAGCTCCAGGATTTGGGCTTCGTGTGCACATTC		2064
Qy	601	AsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyValValThrThr		620
Db	2065	AACCTGCTGAAGATGACAAGCAAGTGAAGCATCTCCGAGGCTGGTGTGGTACCACC		2124
Qy	621	ArgGlnThrArgProaspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrPro		640
Db	2125	AGACAGACACAGGCCGATTCCTTTCCAGGAGAGGTTCCACGGCAGGCAACAAAACCCCA		2184
Qy	641	GluProSerPheProValLysGluProProProValLeuAlaLysProLysLeuAspSer		660
Db	2185	GAGCTTCTTCCCGGTGAAGAGCCCCCTCCAGTTCTTGCCCAAAACCCAACTTGATTTCC		2244
Qy	661	ThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGlnHisGlnLeuGlnAsnPro		680
Db	2245	ACTCAGTTACAACAGCTTCATACCAAGTCTTACTGGAACACACCAATTCGAAACCCCA		2304
Qy	681	ProProSerSerProLysGluPheProPhe***MetThrValLeuAsnSerAsnAlaPro		700
Db	2305	CCTCCTTCATCTCCTAAGGAGTTTCCCTTTTCAGCATGACGTGTTTTCGAACCTCCAATGCTCCC		2364
Qy	701	ProAlaValThrThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeu		720
Db	2365	CCAGCGGTGACAAATCCATGATAGCAGGTGAAGGCTCTTCATCATCAGACGTTACGCTTG		2424
Qy	721	AlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSer		740
Db	2425	GCCCGCCGGAATATTCTTCTCCCTCCACGAACACCCAGCAGCAACTGTGGCCCTTCC		2484
Qy	741	SerSerProValPheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLys		760
Db	2485	AGCTCTCCGGTGTTCATTTTGGCAGCAGCTTCCTCAAACTATTTCAGAGGACGTGAGCAA		2544

Qy	761	GluSerLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIle	780
Db	2545	GAAGGCCTCTTAGTGTCTCACCCCTCTGTGCAAAACCAATCTCCAGGAGGGCTTTTCATC	2604
Qy	781	GlnAsnGluProLeuProGlyProThrGluProThrProProPheThrPheSer	800
Db	2605	CAAAATGAGCGCATCCACCAGGCCAACAGAAACCAACACCACCACCATTCACATTTTCC	2664
Qy	801	IleProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSerProThr	820
Db	2665	ATCCCAGCGGAAACCAGTTTCAGCCCCGCTGTGTGTCCCAATTCTGTCTCTCCTACC	2724
Qy	821	SerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIle	840
Db	2725	AGCCGGATTTCAGAACCCAGTGCGTTTCCTCAGCTGTGTCTGCGCTTCTCTCCCTGCCATC	2784
Qy	841	ProProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeu	860
Db	2785	CCACCACAATAATGCCATGGGCTGCTAGAAGTGACCATCCATGCCATCCACGGGATTA	2844
Qy	861	AlaLysLysAsnThrLysSerProGlnProValAsnAspAspAsnIleArgGluThrLys	880
Db	2845	GCGAAGAAAAATACAAAGTCTCCTCAACAGTGAATGATGATAACATTCGTGAAACTAAG	2904
Qy	881	AsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAsnGln	900
Db	2905	AACGCAGTATTGCGAGACTTGGGGAAAAATAACTTTTCAGTGATGTCAGACCAAACAG	2964
Qy	901	GlnGluTyrlsIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeu	920
Db	2965	CAGGAGTACAAAATTTCAAGCTTTGAGCAGAGGCTGATGAATGAATAGAGTTTCGGTTG	3024
Qy	921	GluArgThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGly	940
Db	3025	GAACGTACTCCTGTTGATGAATCAGATGATGAAATTCAAATGATGAGATCCCCACGGGC	3084
Qy	941	LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer	960
Db	3085	AAGTGATTGTCTCCCATCTTTGACAAGAGACTCAAGCACTTCCGGGTTCACAAGAGCTCT	3144
Qy	961	ProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLys	980
Db	3145	CCAGTCACATTCCACTGCCAAANTTGTGGGATACCTGTGTCCAAGAGTTTACTGGTCAA	3204
Qy	981	AspGlyLysGlnIleSerLysArgAsnGluHisCysLysLysMetArgGluGlyAspGly	1000
Db	3205	GATGGGAAGCAGATTTCTAAGAGAATGAGCGCTGCAAAATGAGCGCAGGAAGAGATGGG	3264
Qy	1001	ThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMet	1020
Db	3265	ACATGCTCTGTGCACATCGAATCCACTACCAAGTATGACGATGGCACTACACCATCATG	3324
Qy	1021	AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeupro	1040
Db	3325	GCAGCAACCCCGAGGGAGATCAGCTGTCTGCGCCACTTGTGTTACAAAGTTTCCGCC	3384
Qy	1041	IleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGlu	1060
Db	3385	ATTTCGAGTCGGCTAACCTCTCTGGTCTGAGTCTCACAGGGGAAGATCCCGAGTGCAGAA	3444
Qy	1061	ArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGly	1080
Db	3445	AGAGACAAAGAGCCCTACGGAAGCGTTTTTCCGACCACCAATTTCTTCGAGGCTCTGGG	3504
Qy	1081	AspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeupro	1100
Db	3505	GATATGATAGTCATCAGGGGCGCTCTCTCGCTGGACTCTAAGTGCAGTGGTTTACCG	3564
Qy	1101	ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAlaSerHisLys	1120
Db	3565	CCCCCGAGCTCACATGGCTACTCAATGGCAACCTGTCTTACAGATGCCCGCCCAAG	3624

Qy	1121	MetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArgAsp	1140
Db	3625	ATGCTGGTCACGGAGACCGGAGTCCACTCTCTGTCTCATTTGACCACCTCACTCAGCGCGAC	3684
Qy	1141	AlaGlyThrTyrlsCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGlu	1160
Db	3695	GCAGGGACCTATAAGTGCGTACCAACAACCGGGCAGAAATCTTTTAGTCTGGAG	3744
Qy	1161	LeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsn	1180
Db	3745	CTCTCTGTAGTACCAAGAGGTGAGNAAGCACCTGTGATCTCTGGAGAAACATACAGAAC	3804
Qy	1181	CysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProPro	1200
Db	3805	TGCGGTGTTCCGAAGGCCACCCGTCGAGAGTCGAGTGC CGCGTGATAGGCATGCCCCCA	3864
Qy	1201	ProValPheTyrrTpLysLysAspAsnGluThrlleProCysThrArgGluArgIleSer	1220
Db	3865	CTCTGTGTTCTACTGGAAAGAACAAATGAGACCATCCCTTGCACGAGAGAGATCAGT	3924
Qy	1221	MethHisGlnAspThrThrGlyTyrlAlaCysLeuLeuIleGlnProAlaLysLysSerAsp	1240
Db	3925	ATGCACACGAGACACACAGGGTATGCTCTTCATTAGCCAGCCAGCAAAATACAGAC	3984
Qy	1241	AlaGlyTrpTyrrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArg	1260
Db	3985	GCTGGATGGTACACGTTGTCAGCCAAGAAATGAAGCCGGCATCGTGTCTGCTGACCTGCCAGG	4044
Qy	1261	LeuAspIleTyrrAlaGlnTrpHisHsglnIleProProPrometSerValargProSer	1280
Db	4045	CTGGATATATACGTCAGTGGCACCATCAGATCCACCGCCCATGCTGTCTCGCGCCAGT	4104
Qy	1281	GlySerArgTyrglySerLeuThrSerLyssglyLeuAspIlePheSerAlaPheSerSer	1300
Db	4105	GGCAGTCCTCAGGATCTCTCACCAGTAAGGACTTGACATATTTTCTGCCTTTCTCTCC	4164
Qy	1301	MetGluserrThrMetValTyrrSerCysserSerArgSerValValGluSerAspGluLeu	1320
Db	4165	ATGGAAGACGAGTGGTGTAFTATCATGCTCTCTCGGAGTGTAGTGGAGAGTGTGAACCTT	4224
RESULT	3	HSM803309	5486 bp mRNA linear PRI 12-JUL-2002
LOCUS		Homo sapiens mRNA; CDNA DKFZp451M023 (from clone DKFz451M023).	
DEFINITION		AL832002	
ACCESSION		AL832002.1	GI:21732542
VERSION			
KEYWORDS		human.	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 5486)	
AUTHORS		Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.	
TITLE		Direct Submission	
JOURNAL		Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY	
COMMENT		Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFz451M023) is available at the R2PD in Berlin. Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@r2pd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/. Location/Qualifiers 1 . 5486 /organism="Homo sapiens" /db_xref="taxon.9606" /clone="DKFz451M023" /tissue_types="cDNA-collection" /clone_lib="451 (synonym: hlcl). Vector pSport1; host DH10B; sites NotI + SalI" /dev_stage="adult"	
FEATURES		source	

polyA_signal		5451.	.5456									
polyA_site		5468										
BASE COUNT	1558 a	1392 c	1260 g	1276 t								
ORIGIN												
Alignment Scores:												
Pred. No.:	2.09e-251	Length:	5486									
Score:	6842.00	Matches:	1309									
Percent Similarity:	99.32%	Conservative:	2									
Best Local Similarity:	99.17%	Mismatches:	9									
Query Match:	99.30%	Indels:	0									
DB:	9	Gaps:	0									
US-09-818-990b-2 (1-1320) x HSW803309 (1-5486)												
Qy	1	MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr	20									
Db	266	ATGCAAGCAGCAGCATAGAGGCTTCTACCTCCATATCTCAGCTTCTAAGAGAGAGCTAT	325									
Qy	21	LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn	40									
Db	326	TTAGCTGAACACGACATCGGGAAACAATGAGAGAGTCGAGCGGAGCCCTCTCTCCAAC	385									
Qy	41	ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlnAspAspLeu	60									
Db	386	CCTTGCCATTTCGGCAGCTCTCTGGGGCGCTGAAGGAGCGGAGGCCAAGATGACCTT	445									
Qy	61	ProAspLeuSerAlaPheLeuSerGlnGluLeuLeuAspGluSerValAsnLeuAlaArg	80									
Db	446	CCAGATCTTTCAGCGCTTTCGAGCCAGAAGAATATGACGAAAGTGCAATTTGGCAAGA	505									
Qy	81	LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu	100									
Db	506	CTGCCATCAATPACGACCTTTGGAGAAGCCAGATGAACCTCAAGTAGAAGACGACTT	565									
Qy	101	SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp	120									
Db	566	TCTCCTGATCAGATGAACACTCACCTTAATTTAAGTTTTCAGGCTTAACCTCTGCGCAGAT	625									
Qy	121	AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGluTyr	140									
Db	626	AACCTCGAAGTCCACAGCTCTAAAGAAGCCCCCAGGAGGCAAAAGGCCACGCTAT	685									
Qy	141	CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu	160									
Db	686	TGTTCTCAACCCAGTCCAAAAAGTATTTTAATTAAGGTCGCCACTTCATTGAAGAG	745									
Qy	161	LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn	180									
Db	746	CTATCTCTCCCTTTTCAATCCACAGCTCCAAAGAGATTAGACCTCGCTGCATAAAC	805									
Qy	181	HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer	200									
Db	806	CACAAGAGTAAACGTGAATCTCAAAACAAAGTTATGCAAGGAAACAGCTCCAGTTCTCA	865									
Qy	201	AspLeuSerGluArgGluArgSerSerValProIleProIleProAlaAspThrArg	220									
Db	866	GATCTGTCAAGAACACGAGAAGATCTCTGTTCCTATCCCTCGGGATACCAGG	925									
Qy	221	AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgGluAlaGluGln	240									
Db	926	GATAATGAAGTGAATCACGCCCTCGAACAGCAGGAAGCCAAAGAGCGTGAAGCGGAGCAG	985									
Qy	241	AlaAlaSerGluAlaAlaGlyLysPheThrThrProGlySerSerProSerSerLeuTyr	260									
Db	986	GCTCCAGTGAAGCGGCTGTGGAGACACTACACAGGGTCTTCCCTTCTCTCTCTAC	1045									
Qy	261	TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal	280									
Db	1046	TATGAAGAACTCTGGGGCAACCTTCCCGGTTTCACTCAAAAGTTTACGAGCAGAGAAGTT	1105									
Qy	281	ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal	300									
Db												
Db	1106	CCAGAAGGAATCCAGTACAGTTGGATTGCATAGTGGTAGAAATTCACCACCTCAAGTA	1165									
Qy	301	ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla	320									
Db	1166	AGTGGTACTGTGAAGCAAGGAGCTTGAANAATCCCAAGATATTCACATCGTCCAGGCA	1225									
Qy	321	GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer	340									
Db	1226	GAAATCTGCACCTACCTGACCATTCGCGAAGCCCTTTCAGAGGACACAGACGCGTATTC	1285									
Qy	341	CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly	360									
Db	1286	TGCTTTTCTTCTTAACATCTATGGGACAGATTCGACTTCTGCTGAGATTTATATAGAAGG	1345									
Qy	361	ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys	380									
Db	1346	GTTTCTTCTTCTGACTCAGAAGGCGCCCTTAACAAGAAGAGATGAATCGAATCCAGAAG	1405									
Qy	381	ProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGln	400									
Db	1406	CCAAATGAGGTGTCATCTCTCCACCTACCTCTGCAGTCACTTCTCCAGCAGTACCCCAA	1465									
Qy	401	AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr	420									
Db	1466	GCCAGCATTTGTTGGCCCAACCTCGTGTGCAACCATCCAGCAGTGTCAAAGCCCAACC	1525									
Qy	421	AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMet	440									
Db	1526	AATTACTTGAGGGATTGGATGGAAACCTATCATTCAGCTCTCTGTTTACAAAGATG	1585									
Qy	441	LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly	460									
Db	1586	CTACAAATTTGTCAGCTCTCTGAGGGTCAGTGGTGTCTTGAATGCAGAGTAAAGGA	1645									
Qy	461	AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAsp	480									
Db	1646	GCTCCATCTCTTAAGTTGAGTGTGTATAGAGAGGAGCTTTAATAGAAGATTCTCCAGT	1705									
Qy	481	PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu	500									
Db	1706	TTTAGGATTTACAGAAAAAACCTCGATCCATGGCAGCCAGGAGAGATTTCACCTTG	1765									
Qy	501	ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys	520									
Db	1766	GTCATTGCTGAGGTTTTCAGAGAGGTTCTGGTGCTTTCACATGCTACTGCAAGCAACAA	1825									
Qy	521	TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn	540									
Db	1826	TACGGCACAGTGTCAAGCATTCACAGCTGCACGCTGAGAGAAATGAGGACCTCAGCAAC	1885									
Qy	541	AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnPro	560									
Db	1886	AACGGGTCTCTTCTACTCAGCCAACTTACCACCACTTGGCAGCTATTGAGCCACAGCCC	1945									
Qy	561	SerProProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGly	580									
Db	1946	TCCCCACCCCACTCAGAGCTCCATCTCTGGAACAACCCCCCAACCACTCGAGGGG	2005									
Qy	581	ValLeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHisPhe	600									
Db	2006	GTTCTGGTGAACCAACAATGAGCCCCGGTCCAGCTCCAGGATTGGCTTCGTGTGCACCTC	2065									
Qy	601	AsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyValValThrThr	620									
Db	2066	GACCTGCTGAAGATGACAAAGGAAGTGAAGCATCTCCCGAGGCTGTGTGTGCACCAACC	2125									
Qy	621	ArgGlnThrArgProAspSer**GlnGluArgPheAsnGlyGlnAlaThrLysThrPro	640									
Db	2126	AGACAGACAGGCCCGGCTTCTTCCAGGAGAGGTTCAACGGCAGGACCAACCAACCCCA	2185									
Qy	641	GluProSerPheProValLysGluProProValLeuAlaLysProLysLeuAspSer	660									
Db	2186	GAGCTTCTTCCCCCGTGAAGAGCCCCCTCCAGTTCTGGCCAAACCCCAACTTGATTCC	2245									

```
Qy 661 ThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGlnHisGlnLeuGlnAsnPro 680
Db 2246 ACTCAGTTACAACAGCTTCATAAACAAGTCTTACTGGAACACACCAATCGCAAAACCCA 2305
Qy 681 ProProSerSerProGlyGluPheProPhe***MetThrValLeuAsnSerAsnAlaPro 700
Db 2306 CCTCCTTCATCTCCCTAAGAGATTCTCTTTCAGCATGACTGTTTGAACATCCAATGCTCCC 2365
Qy 701 ProAlaValThrThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeu 720
Db 2366 CACGCGGTACAAACATCCAGTAAGCAGGTGAAGGCTCTTCATCACACAGCTTCAGCTTG 2425
Qy 721 AlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSer 740
Db 2426 GCCCGGCCGGAAGTATTCTTCCCTCCACGAAACACCCGCGAGCAACTGTGGCCCTTCC 2485
Qy 741 SerSerProValPheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLys 760
Db 2486 AGCTCTCCGGTGTTCACCTTTGAGCAGCACTCTCTCAAACTATTTCAGAGGACAGTGAGCAA 2545
Qy 761 GluSerLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIle 780
Db 2546 GAAAGCCCTTAGTGTCTACCCCTCTGTGCAAAACCAATCTCCAGGAGGCTTTCCATC 2605
Qy 781 GlnAsnGluProLeuProGlyProThrGluProThrProProProPheThrPheSer 800
Db 2606 CAAAATGAGCCACTCCACAGGCCCAACAGAACCAACACACCAACCACTTCACATTTTCC 2665
Qy 801 IleProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSerProThr 820
Db 2666 ATCCCCAGCGGAACACAGTTTCAGCGCCGCTGTGTCTCCCAATTCCTCTCTCTACC 2725
Qy 821 SerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIle 840
Db 2726 AGCCGGATTACAAACCAAGTGTCTCTCAGCTCTGTGTCTGCTCTCTCCCTGCCATC 2785
Qy 841 ProProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeu 860
Db 2786 CCACCCCAAAATGCCATGGGCTGCCTAGAAAGTGCACCATCCATCCAGGGATTA 2845
Qy 861 AlaLysLysAsnThrLysSerProGlnProValAsnAspAsnIleArgGluThrLys 880
Db 2846 GCGAAGAAAAATACAAAGTCTCTCAACACAGTGAATGATGATAACATTCGTGAACACTAAG 2905
Qy 881 AsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAsnGln 900
Db 2906 AACGCGATGATTCGAGACTTGGGGAATAAATAACTTTTCAGTGTGTGACACCAACCCAG 2965
Qy 901 GlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeu 920
Db 2966 CAGGAGTACAAAATTTCAAGCCTTGAGCAGAGGCTGATGAATGAATAGATTTCGCTTG 3025
Qy 921 GluArgThrProValAspGluSerAspGluIleGlnHisAspGluIleProThrGly 940
Db 3026 GAACGTACTCTCTGTTGATGAATCAGATGATGAATAATTCAACTATGATGATGATGATGATG 3085
Qy 941 LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer 960
Db 3086 AAGGTATTGCTCCCATTTTGACAAAGACTCAAGCACTTCCCGGCTCACAAGAGGCTCT 3145
Qy 961 ProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLys 980
Db 3146 CCAGTCACATTCACCTGCAAAATTTCTGGATACCTGTTCCAAAAGTTTACTGGTTCAAA 3205
Qy 981 AspGlyLysGlnIleSerLysArgAsnGluHisCysLysLysMetArgArgGluGlyAspGly 1000
Db 3206 GATGGGAAGCAGATTTCTTAAGAGAATAGCACTGCAAAATGAGCGGAGAGAGATGGG 3265
Qy 1001 ThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMet 1020
Db 3266 ACATGCTCTTCACATTTGAATCCACTACCACTGATGATGATGATGATGATGATGATGATGATG 3325
```

```
Qy 1021 AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro 1040
Db 3326 GCAGCAACCCAGGAGGAGAAATCAGCTGTCTGSCCACTTGATGGTACAAAGTTTGCC 3385
Qy 1041 IleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGlnGlu 1060
Db 3386 ATTGCACTCGCTTAACCTCTGCTGGTCAGTCTCACAGGGAAGATCCCGAGTGAAGAA 3445
Qy 1061 ArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaProGly 1080
Db 3446 AGAGCAAAAGAGCCCTACAGAACGCTTTTCCAGCCACATTTCTCGAGGCTCCTGGG 3505
Qy 1081 AspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeuPro 1100
Db 3506 GATATGGTAGCTCATGAGGGCGCTCTGTGCTGGACTGTAAAGTGAGTGGTTTACCG 3565
Qy 1101 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHisLys 1120
Db 3566 CCCCAGGAGCTGACATGGCTACTCAATGGCAACGCTGTGTCTTACCATGCTCCACAAAG 3625
Qy 1121 MetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnArgAsp 1140
Db 3626 ATGCTGTGTCAGGAGACCGGAGTCCACTCTCTCTCATTTGACCCACTCAGCGCCAG 3685
Qy 1141 AlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGlu 1160
Db 3686 GCAGGGACCTATAAGTGCATCGCTACCAACAAAACCCGGCAGAAATCTTTTAGTCTGGAG 3745
Qy 1161 LeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsn 1180
Db 3746 CTCTCTGTAGTAGCAAAAGAGGTGAAGAAAGCAGCTGTGTATCTCTGGAGAAACTACAGA 3805
Qy 1181 CysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProPro 1200
Db 3806 TCGCGTGTTCGCAAGGCCACCCGCTGAGCTGGAGTCCGCGGTGATAGGCGATGCCCCCA 3865
Qy 1201 ProValPheTyrTrpLysLysAsnGluThrIleProCysThrArgGluArgLysSer 1220
Db 3866 CCTGTGTCTTACTGGAAGAAAGACAATGAGACCATCCCTTCACACAGAGAGAGATCAGT 3925
Qy 1221 MetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAsp 1240
Db 3926 ATGCACCAAGGACAAACAGGATGCTCTGCTCTCATTTACGCCAGCCCAAGAAATACAG 3985
Qy 1241 AlaGlyTyrTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArg 1260
Db 3986 CTGATGGTACAGTGTGTCAGCAAGAAATGAAGCCGCACTGCTGTGCTGCTGCCAGG 4045
Qy 1261 LeuAspIleTyrAlaGlnTrpHisGlnIleProProProMetSerValArgProSer 1280
Db 4046 CTGGATATATACGCTCAGTGCACCATCAGATCCACGCCCATGCTGTCTCGGCCCATG 4105
Qy 1281 GlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSerSer 1300
Db 4106 GGCAGTCTGCTACGATCTCTCACCAGTAAAGGACTTGACATATTTTCTGCTCTTCTCC 4165
Qy 1301 MetGluSerThrMetValTyrSerCysSerSerArgSerValGluSerAspGluLeu 1320
Db 4166 ATGGAAGACCATGGTGTATTTCATGCTCTCTCGAGTGTAGTGGAGATGAACCTT 4225
```

```
RESULT 4
HSM803687 HSM803687 4683 bp mRNA linear PRI 10-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp451K027 (from clone DKFZp451K027).
DEFINITION AL832379
ACCESSION AL832379.1 GI:21732942
VERSION human.
KEYWORDS SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4683)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.
```

TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Sequenced by BMFZ (Biomedical Research Center at the Charite, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp451k027) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

source Location/Qualifiers

1..4683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp451k027"
/tissue_type="cDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
4576..4581
polyA_signal

polyA_site 4594

BASE COUNT 1376 a 1147 c 1048 g 1112 t
ORIGIN

Alignment Scores:

Pred. No.: 9,59e-194 Length: 4683
Score: 5322.00 Matches: 1013
Percent Similarity: 99.41% Conservative: 0
Best Local Similarity: 99.41% Mismatches: 6
Query Match: 77.24% Indels: 0
DB: 9 Gaps: 0

US-09-818-990B-2 (1-1320) x HSM803687 (1-4683)

QY 302 TrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAlaGly 321
DB 293 TGGTACTGTGAAGGCAAGGAGCTTGAATAATCCACATATTCACATCGTCCAGGACGA 352
QY 322 AsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSerCys 341
DB 353 AATCTGCACCTACACGACATTCGCGAAGCCCTTTGAAGGACACAGACGCTATTCTCTGC 412
QY 342 PheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGlyVal 361
DB 413 TTTCTTCTAACATCTATAGGACAGATTCGACTTCTCGAGATTTATATAGAAGGGTT 472
QY 362 SerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLysPro 381
DB 473 TCTTCTTCTGACTCAGAAGGCGACCCCTAACAAAGGAGAGATGAATCGAATCCAGAAGCCA 532
QY 382 AsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGluAla 401
DB 533 AATGAGTGTGATCTCTCCACTACCTCTGCAGTCATCTCTCCAGCAGTACCCCAAGCC 592
QY 402 GlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThrAsn 421
DB 593 CAGCATTTGGTGGCCCAACCTCTGTGGCAACCATCCAGCAGTGTCAAGCCCCCACCAT 652
QY 422 TyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMetLeu 441
DB 653 TACTTGCAGGATTTGGATGGAAACCTATCATTCAGCTCCTGTGTTTACAAAGATGCTA 712
QY 442 GlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGlyAla 461
DB 713 CAAATTTGTACAGTCTTGAGGGTTCAGCTGGTTCTTGAATGCAGAGTAAAGAGAGCT 772
QY 462 ProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAspPhe 481
DB 773 CCATCTCTTAAGGTGTAGTGTAGAGAGGGAGCTTTAATAGAAGATTCCTCCAGATTT 832
QY 482 ArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeuVal 501

DB 833 AGGATTTTACAGAAAAACCTCGATCCATGCAGAGCCAGAGGAGATTTCACCTTGGTC 892
QY 502 IleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLysTyr 521
DB 893 ATTGCTGAGTGTGTTCAGAGAAGATTCTGGGTGCTTACATGTACTGCAAGCAACAAATAC 952
QY 522 GlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsnAsn 541
DB 953 GGCACAGTGTCAAGCATTTGCACAGCTGCAGTGCAGAGAAATGAGGACCTTCAGCAACAC 1012
QY 542 GlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnProSer 561
DB 1013 GGGTCTCTTCACTCAGCCAACCTTACCACCAACCTGCAGCTATTGAGCCACAGCCCTCC 1072
QY 562 ProProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGlyVal 581
DB 1073 CCACCCCACTCAGAGCCTCCATCTGTGGAACAACCCCAACCAACCTCGAGGGGTT 1132
QY 582 LeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsn 601
DB 1133 CTGGTGAACCAATGAGCCCGCTCCAGCTCCAGGATTGGGCTTCGTGTGCACCTTCAAC 1192
QY 602 LeuProGluAspLysGlySerGluAlaSerSerGluAlaGlyValValThrThrArg 621
DB 1193 CTGCCTGAAGTGAACAAGGAAGTGAAGCATCTCCGAGGCTGGTGGTGACCAACCA 1252
QY 622 GlnThrArgProAspSer***GlnGluArgPheAsnGlyGlnAlaThrLysThrProGlu 641
DB 1253 CAGACCAAGGCCGATTTCTTCCAGGAGAGGTTCACGACGAGGCAACAAACCCACAG 1312
QY 642 ProSerPheProValLysGluProProValLeuAlaLysProLysLeuAspSerThr 661
DB 1313 CTTTCTTCCCGGTGAAGAGCCCTCCAGTCTCGCCAAACCAACCTTGATTCCACT 1372
QY 662 GlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGlnGlnHisGlnLeuGlnAsnPro 681
DB 1373 CAGTTACACAGCTTCATAACCAAGTCTTACTGGAACAACCAACCAATTCGAAAACCCACCT 1432
QY 682 ProSerSerProLysGluPheProPhe***MetThrValLeuAsnSerAsnAlaProPro 701
DB 1433 CTTTCACTCTCTAAGAGAGTTTCTTTCAGCATGACTGTTTGAACCTCCAATGCTCCCA 1492
QY 702 AlaValThrThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeuAla 721
DB 1493 GCGGTGACACATCCAGTAAGCAGTGAAGCTCTCTTCAATCAGACAGCTTCAGCTTGGCC 1552
QY 722 ArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProSerSer 741
DB 1553 CGGCCGAAGTATTCTTCTCCCTCCACGAACACCAACCACTGTGGCCCTTCCAGC 1612
QY 742 SerProValPheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerLysGlu 761
DB 1613 TCTCCGGTGTTCACCTTTGAGCAGACTCTCCAAACTATTCAGAGGACAGTGAAGAA 1672
QY 762 SerLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIleGln 781
DB 1673 AGCCTCTTAGTGTCTACCCCTCTGTGCAACCAAACTCCAGGAGGCTTTCATCCAA 1732
QY 782 AsnGluProLeuProProGlyProThrGluProThrProProProPheThrPheSerIle 801
DB 1733 AATCAGCCACTCCCAAGGCCCCAACCAACCAACCAACCACTTCCATTTTCCATC 1792
QY 802 ProSerGlyAsnGlnPheGlnProArgCysValSerProIleProValSerProThrSer 821
DB 1793 CCCAGCGAAACCAAGTTTCAGCCCCGTGTGTGCCCAATTCCTGTCTCTCCACACAGC 1852
QY 822 ArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAlaIlePro 841
DB 1853 CGGATTCAGAACCCAGTGGCTTCTCTCAGCTCTGTTCTGCGCTTCTCTCCCTGCCATCCA 1912
QY 842 ProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerGlnGlyLeuAla 861

GPSGAAEGGGQDDLPDLSAFLSQBELDESVNARLAINYDPLEKADETQARRKRLSP
DOMKSPNSFENFQDNPRPTSSKESQAEKRPQYQSETQSKKFKVFNKAADFIEE
LSLEKSHSKRPRACKNKKLESONKVMQENSSFSDLSERRERSVPTIPAD
TRNEVNHLEQOEAKRREAQAASEAAGDTPPGSSPSLYVEEPLGPPRFTQKLR
SRVPEGRVQLDCIVVGPVVRNVCBKELENSPDHIHQVQGNLHSLUTIAEAE
DTGRYSCFANSYGTSTAEIIEGVSSDSSEGDKNEMNRIOQNPNEVSSPTTSA
VTPAVFOAHLVAQPRVATIQOCQPTNYLQGLDKPIIAAPVFTKMLQNLASSEGO
LVVFECRVGAQSPKVEWYREGTLIEDSPDFRLQKPRSMABPGKDFNFNLLVVE*

BASE COUNT 638 a 501 c 487 g 501 t

Alignment Scores:

Pred. No.: 3, 29e-89 Length: 2127
Score: 2560.00 Matches: 494
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.16% Indels: 0
DB: 9 Gaps: 0

US-09-818-990B-2 (1-1320) x AK027343 (1-2127)

Qy 1 MetcInAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 231 ATGCAAGACGACAGATAGAGCTTCTACTTCCATATCTCAGCTTCTAAAGAGAGAGCTAT 290
Qy 21 LeuAlaGluThrArgHisArgGlyAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 291 TTACTGAACACGACATCGGGGAAACAATGAGAGAGTGCAGCGGAGCCCTCTCCCAAC 350
Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeu 60
Db 351 CCTTGCCATTTGCGCAGTCTCTCTGGGGCGCTCAAGGAGCGGAGCCAGATGACCTT 410
Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 411 CCAGATCTTTCAGCCTTCTGAGCCAGAGAAGATTAGACGAAAGTGCAATTTGGCAAGA 470
Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Db 471 CTGCCATCAATTACGACCTTTTGGAGAGGACAGATGAATCTCAAGTACGAAACGACTT 530
Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Db 531 TCTCCTGATCAGATGAACACTCACCTAAATTTAAGTTTTCAGCCTAACTTCTGCCAGAT 590
Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Db 591 AACCTCGAAGTCCACACGCTCTAAAGAAAGCCCGGAGGAGGCAAAAGGCCACAGTAT 650
Qy 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Db 651 TGTCTGNAACCCAGTCCAAAAAGATATTTTAATRAGGCTGCGACTTCATTGAAGAG 710
Qy 161 LeuSerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsn 180
Db 711 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAGAGATTAGACCTCGTGCCTGCAAAAAC 770
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSerPheSer 200
Db 771 CACAAGAGTAACCTGGAATCTCAAAACAAGATTATGCAAGGAAAAACAGCTCCAGTTCTCA 830
Qy 201 AspLeuSerGluArgArgGluArgSerValProIleProIleProAlaAspThrArg 220
Db 831 GATCTGTGAGAAAGACGAGAAAGATCTTCTGTTCCATCCCTATCCCTGCGGATACCAAG 890
Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
Db 891 GATAATGAAGTGAATCAGCCCTCGAACAGCAGGAGCAAGAGCGCTGAAGCGGAGCAG 950
Qy 241 AlaAlaSerGluAlaAlaGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
Db 951 GCTGCCAGTGGAGGGGTGGTGGAGACACTACACAGGGTCTTCCCTCTCATCTCTGTAC 1010

Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
Db 1011 TATCAAGAACCTCTGGGSCAACCTCCCGGTTCACCTCAAAAGTTTACGAGCAGAGATT 1070
Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
Db 1071 CCAGAAGAACTCGAGTACAGTTGGATTGCTAGTAGTAGGAATTCACCACTCAAGTA 1130
Qy 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
Db 1131 AGTGGTACTGTGAAGCAAGGAGCTTGAANAATTTCCCAATATTCACATCGTCAGGCA 1190
Qy 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
Db 1191 GGAATCTGCACCTCCTGACCATTTGCGAAGCCTTTGAAGAGGACACAGAGCGTATTC 1250
Qy 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
Db 1251 TGCCTTCTTCTTAACATCTATGGGACAGATTCGAGTCTCTGCTGAGATTTATATAGAAGG 1310
Qy 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
Db 1311 GTTCTTCTTCTGACTCAGAAGCGCCCTTAACAAGGAAGAGATGAATCGAATCCAGAAG 1370
Qy 381 ProAsnGluValSerSerProThrThrSerAlaValIleProProAlaValProGln 400
Db 1371 CCAATGAGGTGTCTATCTCTCCCACTACCTCTGCTGCTGCTCTCCAGCAGTACCCCA 1430
Qy 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
Db 1431 GCCCAGCATTTGTTGGCCCAACCTCGTGTGGCAACCACTCCAGCAGTGTCCAGAGCCCC 1490
Qy 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaProValPheThrLysMet 440
Db 1491 AATTACTTGCAGGATTGGATGGAAAACCTATCATGTGAGCTCTCTGTTTACAAAGATG 1550
Qy 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
Db 1551 CTACAAAATTTGTGCTGCTCTGAGGGTCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTG 1610
Qy 461 AlaProSerProLysValGluTyrTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
Db 1611 GCTCCATCTCTTAAGTTGAGTGTGATAGAGAAGGAGCTTTAATAGAAGATTTCTCCAGAT 1670
Qy 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluPro 494
Db 1671 TTTAGGATTTTACAGAAAAAACCTCGATCCATGGCAGAGCCA 1712
RESULT 6
AF464873 5773 bp mRNA linear PRI 11-FEB-2002
LOCUS AF464873 Homo sapiens myoneurin (MYN) mRNA, complete cds.
DEFINITION AF464873
ACCESSION AF464873
VERSION AF464873.1 GI:18448934
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE
1 (bases 1 to 5773)
AUTHORS Lockwood, S.K. and Sims, K.B.
TITLE A protein with two immunoglobulin-like domains interacts with the
Norrie disease gene product norrin
JOURNAL Am. J. Hum. Genet. 61, A158 (1997)
REFERENCE 2 (bases 1 to 5773)
AUTHORS Lockwood, S.K.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-2001) Neuroscience Center, Massachusetts General
Hospital, 149 13th Street, Charlestown, MA 02129, USA
FEATURES
Location/Qualifiers
1..5773
/db_xref="taxon:9606"

Db 1409 TCAATCTCCAAAGACAGGGGTGACCACAGCTGTGATT----- 1444
QY 403 HisLeuValAlaGlnProArgValAlaThrIleGlnClnCysGlnSerProThrAsnTyr 422
Db 1445 -----CAACCACTGTCTGCTGTGCAACAGGTTACAGATCCAACTTCATAT 1492
QY 423 LeuGlnGlyLeuAspGly---LysProIleIleAlaAlaProValPheThrLysMetLeu 441
Db 1493 CTCTGCCGACCTGATGGAACCACTACTGCCTACTTCTCTCTGTTTTTACAAGGAACGTG 1552
QY 442 GlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGlyAla 461
Db 1553 CAAACACACAGCGTGGCGGAAGCCAGGTGGTGTCTGGAGTGGCGGGTCCGTGGGGCA 1612
QY 462 ProSerProLysValGluThrPtyArgGluGlyThrLeuIleGluAspSerProAspPhe 481
Db 1613 CCCCTCTCAGTCCAGTGGTTCGCAAGGAGGTGAATCCAAAGACTCTCCAGATTTC 1672
QY 482 ArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeuVal 501
Db 1673 CGAATTTCTACAGAAAACCTAGATACAGCTGAACCTGAGGAGATTTCACCCCTAGTT 1732
QY 502 IleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLysTyr 521
Db 1733 ATCGCTGAGACTTTCCTGAAGATGCAGGATCTTTACATGTTTCAGCAAGAAATGATTAT 1792
QY 522 GlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsnAsn 541
Db 1793 GGATCAGCAACAGCACTGCCCTGCTGTCACCTCAGCCCAAC-----ACTGAAAAC 1846
QY 542 GlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaIleGluProGlnProSer 561
Db 1847 TGTAGTTACGAGTCAATGGAGNATCCACATGACCATGACCATTCCACACTTCCACCTCCC 1906
QY 562 ProProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGlyVal 581
Db 1907 CTTCCAATCTTGGAGACAGTTCTCTTGGAGTTGGCTTCAAAGAAACCATCTGAGATCCAG 1966
QY 582 LeuValAsnHisAsnGluProArgSerSerArgIleGlyLeuArgValHisPheAsn 601
Db 1967 CAGGTGAAC---AACCTGAGTTAGCCCTGAGCAGGCGACCCCTCAAATGCAATTCAT 2023
QY 602 LeuProGluAspLysGlySerGluAlaSerSerGluAlaGlyValValThrArg 621
Db 2024 GCTGCTGAGAGGAA-----ACGAGCGAGTCCATCCACGCCGT 2062
QY 622 GlnThrArgProAspSer**GlnGluArgPheAsnGlyGlnAlaThrLysThr----- 639
Db 2063 -----GGAGTAAATGGACTGATTAAACGGCAAGCTAACAGTAATAATCT 2107
QY 640 ---ProGluProSerPhe-----ProValLysGluProProValLeuAlaLys 655
Db 2108 CTTTCAACACCACTGCTCGCTTCCACCCACTTAAGAGCCACCACTCTGCTGTGCCAA 2167
QY 656 ProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGlnHis 675
Db 2168 CCAAACTA----- 2176
QY 676 GlnLeuGlnAsnProProProSerSerProLysGluPheProPhe**MetThrValLeu 695
Db 2176 ----- 2176
QY 696 AsnSerAsnAlaProProAlaValThrThrSer**LysGlnValLysAlaProSerSer 715
Db 2176 ----- 2176
QY 716 GlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAla 735
Db 2176 ----- 2176
QY 736 ThrValAlaProSerSerSerProValPheThrLeuSerSerThrProGlnThrIleGln 755

Db 2176 ----- 2176
QY 756 ArgThrValSerLysGluSerLeuLeuValSerHisProSerValGlnThrLysSerPro 775
Db 2176 ----- 2176
QY 776 GlyGlyLeuSerIleGlnAsnGluProLeuProProGlyProThrGluProThrProPro 795
Db 2176 ----- 2176
QY 796 ProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysValSerProIle 815
Db 2176 ----- 2176
QY 816 ProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSerSerValLeuPro 835
Db 2176 ----- 2176
QY 836 SerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMet 855
Db 2176 ----- 2176
QY 856 ProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAspAsn 875
Db 2177 -----GGATTTCCAAAGAGGCCAGTAGAACTGCTAGAAATAGCTCCGATGAGAA 2227
QY 876 IleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAsp 895
Db 2228 ATTCAAGGCACAAAGGATGCTGTTATTCAAGACCTGGAACGAAACTTCGCTTCAAGGAG 2287
QY 896 ValArgProAsn----- 899
Db 2288 GACCTCTGNACAAATGCCAGCCGAGGTTAAACATACGAAAGAAAGATGGCTCTGCACCTG 2347
QY 900 -----Gln 900
Db 2348 CTAGGTGCTGACAGTGCAACTGCTTTAAATATTTCAGGAGCCAGAGGAAACAGACTAAT 2407
QY 901 GlnGluThrLysLysSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeu 920
Db 2408 CAGGAATACAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATAGAGTAGTACAGGCTA 2467
QY 921 GluArgThrProValAspGluSerAspGluIleGlnHisAspGluIleProThrGly 940
Db 2468 GAAAGTCTCTCTGGATGAATCAGGTGATGAAGTTCAGTATGAGATGTCCTGTGGAA 2527
QY 941 LysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySer 960
Db 2528 AATGGAATGGCACCATTCTTTGAGATGAAGCTGAAACATTTACAAGATCTTTTGAAGGAATG 2587
QY 961 ProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLys 980
Db 2588 CCAGTAACCTTTCATGTAGATGGCTGGAAATCCAAAGCCAAAGATCTATTGTTTAAA 2647
QY 981 AspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAspGly 1000
Db 2648 GATGGGAAGCAGATCTCTCCAAAGAGTGATCACTACACCATTTCAAAGAGATCTCGATGG 2707
QY 1001 ThrCysSerLeuHisIleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMet 1020
Db 2708 ACCTGCTCCCTCCATCCACAGGCTCCACCTAGATGATGATGGATTTATACAAATTTATG 2767
QY 1021 AlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuPro 1040
Db 2768 GCTGCAAAACCTCAGGCGCGCATCAGTTGTACTGGAGCGCTAATGGTACAGGCTGTCAAC 2827
QY 1041 IleArgSerArgLeuThrSerAlaGlyGlnSerHis-----ArgGly 1054
Db 2828 CAAAGAGTTCGAAGTCCCCGCTCTCCCTCAGGCCATCTCATGTCAGAAAGCCCTCGTCT 2887
QY 1055 ArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProHis 1074
Db 2888 AGATCAAGGGCAGCTGGAGACGAAATGAACCAATTTCAAGGAGCGATTCTTTCAAGACCTCAC 2947

```
Qy 1075 PheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCys 1094
|||||
Db 2948 TTCTTGCAGGCTCTGGAGATCTGACTTTCAGGAAGGAAACTCTGCAGAAATGGACTGC 3007
|||||
Qy 1095 LysValSerGlyLeuProProGlyLeuThrTrpLeuLeuAsnGlyGlnProValLeu 1114
|||||
Db 3008 AAAGTCAGTGGTTACCAACCCAGATCTAAGCTGGCAAGCTAGATGGAAGACCCGTCAGC 3067
|||||
Qy 1115 ProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAsp 1134
|||||
Db 3068 CCTGACAGTGTCTACAGATGCTGGTGGTGAGAACGGGTGCATCTCTCATATGATAGAG 3127
|||||
Qy 1135 ProLeuThrGlnArgAspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGln 1154
|||||
Db 3128 CCAGTCAGTCACGTGATCGCGCATCTACACATGTATAGTACCAACCGAGCAGCAG 3187
|||||
Qy 1155 AsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAlaProValIle 1174
|||||
Db 3188 AACTCATTTACGCTGGAGCTTGTGTTGCTGCTAAAGAACACACAAACCCCTGTGTTT 3247
|||||
Qy 1175 LeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArg 1194
|||||
Db 3248 ATTGAGAAAGCTCCAAACACAGAGAGTGTCTGATGGGTACCCAGTGCAGCTGGAATGTCGT 3307
|||||
Qy 1195 ValIleGlyMetProProValPheTrpLysLysAspAsnGluThrIleProCys 1214
|||||
Db 3308 GTATTGGAGTGCACCACTCAGATATTTTGGGAAGAAAGAAATGAATCACTCACTCAC 3367
|||||
Qy 1215 ThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGln 1234
|||||
Db 3368 AGCACTGACCGAGTGAGCATGCACAGCAGCAACCAACCGCTACATCTGCTGCTCATTCAG 3427
|||||
Qy 1235 ProAlaLysLysSerAspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIle 1254
|||||
Db 3428 GGAGCCACAAAGAGATGCTGGTGGTGATGCTGTGACGCAAGATGAAGCAGGGATT 3487
|||||
Qy 1255 ValSerCysThrAlaArgLeuAspIleTyr-----GlnIleProProMetSerVa 1277
|||||
Db 3548 GGAGAACAAATACCAACCCAGTGGCATTCAGCATGCACGACCAAGCCCAAAAAAGT 3607
|||||
Qy 1277 lArgProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAl 1297
|||||
Db 3608 ACGGCCCTCAGCAGTGCCTATGACGACCTTTCGGACCAAGGACTAGACATCAAGCAGC 3667
|||||
Qy 1297 aPheSerSerMetGluSerThrMetValTyrSerCysSerSerArgSerValValGluSe 1317
|||||
Db 3668 GTTCCAACT---GAGGCCAACCCATCTCACCTGCACACTGAATACTGCTTGTGTAGAAAG 3724
|||||
Qy 1317 rAspGluLeu 1320
|||||
Db 3725 TGAGGACCTG 3734
|||||
RESULT 7
AK095512 2920 bp mRNA linear PRI 15-JUL-2002
LOCUS AK095512
DEFINITION Homo sapiens cDNA FLJ38193 fis, clone FCBBF1000280.
ACCESSION AK095512
VERSION AK095512.1 GI:21754783
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens fetal brain cDNA to mRNA, clone_lib:FCBBF1
clone:FCBBF1000280.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A.,
```

```

Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K.,
Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y.,
Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K.,
Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A.,
Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
REFERENCE 2 (bases 1 to 2920)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL
COMMENT Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatori, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source Location/Qualifiers
1..2920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FCBBF1000280"
/tissue_type="brain"
/clone_lib="FCBBF1"
/dev_stage="fetal"
/note="cloning vector: pME18SFL3"
BASE COUNT 839 a 736 c 718 g 627 t
ORIGIN
Alignment Scores: Length: 2920
Pred. No.: 6.99e-71 Matches: 475
Score: 2081.50
Percent Similarity: 51.79% Conservative: 132
Best Local Similarity: 40.53% Mismatches: 249
Query Match: 30.21% Indels: 318
DB: 9 Gaps: 23
US-09-818-990B-2 (1-1320) x AK095512 (1-2920)
Qy 228 LeuGluGlnGlnGlnAlaLysArgArgGluAlaGluGlnAlaLysSerGluAlaAlaGly 247
|||||
Db 13 CTGCGCCGCCAGGAGCTTCCCGGAGCCGCGGGCCAGGACCTCTCGAGCGCAGCAG 72
|||||
Qy 248 GlyAspThrThrPro---GlySerSerProSerSerLeuTyrTyrGluGluProLeu--- 265
|||||
Db 73 GCAAGGGCGCGCCTTTTGGCAGCGCGCTGCCAGTTAAATTT-----TCTCCGTTGCAT 126
|||||
Qy 266 -----GlyGlnProProArgPheThrGlnLysLeuArgSer 277
|||||
Db 127 TCTGCAACAGAGATCGCATTCAGAGCCCAAGTTGGAAGCTGGGAATAAGCG-AGATCT 185
|||||
Qy 278 ArgGluValProGluGlyThrArgValGlnLeuAspCysIleValValGly----- 294
|||||
Db 186 GAAAGACCCCAAGAGAGCT-----TGTGGGAGTGGGGCAGACGGC 230
|||||
Qy 295 ---IleProProGlnValArgTrpTyrCysGluGlyLysGluLeuGluAsnSerPro 313
|||||
Db 231 GTTTTGGCCCA-----ACTGA-TGTTTCTGTGAAGGAAAGAACTGCACACACTCCT 283
|||||
Qy 314 AspIleHisIle---ValGlnAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPhe 332
|||||
Db 284 GATATTCAATTCCTACTGTGAGGGAGGGGACCTCATACCTGATCATAGCAGAGCCCTT 343
|||||
Qy 333 GluGluAspThrGlyArgTyrSerCysPheAlaSerAsnIleTyrGlyThrAspSerThr 352
|||||
Db 344 GAGGACGACAGAGTCCCTACACCTGTTGCTACGAAATCCACCGGCTCAGACACAA 403
|||||
Qy 353 SerAlaGluIleTyrIleGluGlyValSerSerSerGluGlyAsp----- 369
|||||
```

Db 404 TCTGCTGAGGTGTTATTTGAAGTGGCAGTTCAACAGATTCTCACAGTGAAGTTAGCT 463
QY 370 -----ProAsnLysGluGluMetAsnArgIleGlnLysProAsnGlu----- 383
Db 464 TTCAATCAAGAGCTGGAGCTGACCAAGCTCAAAAGAAACAACACTTCTGTTCTCTTG 523
QY 384 -----ValSerSerProThr-----ThrSerAlaValIleProProAlaVal 398
Db 524 ACAATAGGATCAATCTCCAAAGACAGGGGTGACCACAGCTCTGATT----- 571
QY 399 ProGlnAlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSer 418
Db 572 -----CAACCACCTGCTCTCCCTGTGCAACAGGTTCACAGT 607
QY 419 ProThrAsnTyrLeuGlnGlyLeuAspGly---LysProIleIleAlaAlaProValPhe 437
Db 608 CCAACTTCATATCTTCGCGACCTGATGGAAACCACACTACTGCTCTCTCTCTGTTTT 667
QY 438 ThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArg 457
Db 668 ACAAGGAAGCTGCAAAACACAGCGTGGGGAAGCCAGGTGGTCTGAGTGCCTGG 727
QY 458 ValLysGlyAlaProSerProLysValGluTyrTyrArgGluGlyThrLeuIleGluAsp 477
Db 728 GTCCGTGGGCGACCCCTCTGCAAGTCCAGTGTGTCGCAAGGAGTGAATCCCAAGAC 787
QY 478 SerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIle 497
Db 788 TCCTCCAGATTCCGAATCTTACAGAAAAACCTAGATCTACAGCTGAACCTGAGGAGATT 847
QY 498 CysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAla 517
Db 848 TGCACCTAGTTATCGGAGACTTCCCTGGAAGATGCAGGGATCTTACATGTTAGCA 907
QY 518 SerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAsp 537
Db 908 AGAAATGATTATGGATCAGCAACAGCAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCT 964
QY 538 LeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaAlaIleGlu 557
Db 965 ---ACGAAACCTGTAGTTACGAGTCAATGGGAGATCCAAACATGACCACTTCCAAAC 1021
QY 558 ProGlnProSerProHisSerGluProProSerValGluGlnProProLysProLys 577
Db 1022 TTTCCACCTCCCTCCAACTTTGGAGACAAGTTCTCTGGAGTTGGCTTCAAGAAACCA 1081
QY 578 LeuGluGlyValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArg 597
Db 1082 TCTGAGATCCAGCAGGTGAAC---AACCCCTGAGTTAGGCTGAGCAGGCGAGCCCTCAA 1138
QY 598 ValHisPheAsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyVal 617
Db 1139 ATCAATTCATGCTGCTGAGAGGAACCAAC-----GGAGTC 1177
QY 618 ValThrThrArgGlnThrArgProAspSer***GlnGluArgPheAsnGlyGlnAlaThr 637
Db 1178 CATCCCGAGCGT-----GGAGTAAATGAGTGTATACGCCAAGCTTAAC 1222
QY 638 LysThr-----ProGluProSerPhe-----ProValLysGluProProPro 651
Db 1223 AGTAATAAATCTCTTCAACACACAGCTGCTGCTGCTTTCACCCCACTAAGGAGCCACCT 1282
QY 652 ValIleAlaLysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeu 671
Db 1283 CTGCTGCCAAACCAAACTA----- 1303
QY 672 LeuGluGlnHisGlnLeuGlnAsnProProProSerProLysGluPheProPhe*** 691
Db 1303 ----- 1303
QY 692 MetThrValLeuAsnSerAsnAlaProProAlaValThrThrSer***LysGlnValLys 711

Db 1303 ----- 1303
QY 712 AlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSerThrAsn 731
Db 1303 ----- 1303
QY 732 ThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSerSerThrPro 751
Db 1303 ----- 1303
QY 752 GlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHisProSerValGln 771
Db 1303 ----- 1303
QY 772 ThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProGlyProThrGlu 791
Db 1303 ----- 1303
QY 792 ProThrProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCys 811
Db 1303 ----- 1303
QY 812 ValSerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSer 831
Db 1303 ----- 1303
QY 832 SerValLeuProSerLeuProAlaIleProProThrAsnAlaMet***LeuProArgSer 851
Db 1303 ----- 1303
QY 852 AlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProVal 871
Db 1304 -----GGATTTCAAAGAGCCAGTACAGTCTAGTAATGACC 1342
QY 872 AsnAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIle 891
Db 1343 TCCGATGAGGAAATTCAGGCACAAAGGATGCTGTTATTCAAGACCTGGAACGAAACTT 1402
QY 892 ThrPheSerAspValArgProAsn----- 899
Db 1403 CGTTTCAAGGAGGAGCTCTCTCAACATGGCCAGCCGAGTTAATACAGAAAGAAGATG 1462
QY 899 ----- 899
Db 1463 GCTCGTGGAGTGTAGTGTGACAGTGCACATGCTCTTTAATATTACGAGGCCAGAGAG 1522
QY 900 -----GlnGlnGluTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIle 916
Db 1523 GAAACAGCTAATCAGGAATACAAAGTCTCCAGCTGTGCAACAGAGACTCATCAGTGAATA 1582
QY 917 GluPheArgLeuGluArgThrProValAspLysSerAspAspGluIleGlnHisAspGlu 936
Db 1583 GAGTACAGGCTAGAAAGCTCTCTGTGATGAATCAGCTGATGAAGTTCAGTATGGAGAT 1642
QY 937 IleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgVal 956
Db 1643 GTGCTGTGGAATAATGGAATGGACCATCTTTGAGATGAAGCTGAACATTTACAAGATC 1702
QY 957 ThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysVal 976
Db 1703 TTTGAGGAAATCCAGTAACTTTTCACATGATAGTGGCTGGAATCCAAAGCAAGATC 1762
QY 977 TyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArg 996
Db 1763 TATTGGTTTAAAGATGGGAAGCAGATCTCTCCAAAGAGTATCATCACTACACATTTCAAGA 1822
QY 997 GluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsn 1016
Db 1823 GATCTCATGGAGCTGCTCTCCCTCCATACACACCTCCACCTAGATGATGATGGAAT 1882
QY 1017 TyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetVal 1036
Db 1883 TATACAATTATGGTGCAAACCCCTCAGGCGCGCATCAGTTCTACTGGACGCTAATGTA 1942

Db 527 TCAGCTCCGAAACAGGAGTGACACAGCTGTGTGATTGAGCCCTTATCTGCTGTGCTCAAA 586
Qy 415 GlnCysGlnSerProThrAsnTyrLeuGlnGlyLeuAspGlyLysProIle---IleAla 433
Db 587 CAGGCTCACAGCGTACTTATATCTCTGGGACCTGACGGAACCCACCATGGGTGCTT 646
Qy 434 AlaProValPheThrLysMetLeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValVal 453
Db 647 CTTCCTGTTTTACTAAGGAACATACACAACACAGCAGCCCTCCGAGGCGCAGGTGGTG 706
Qy 454 PheGluCysArgValLysGlyAlaProSerProLysValGluTyrPyrArgGluGlyThr 473
Db 707 CTGAGTGCAGAGTCGAGGGGACCCCGCTGACGATCCAGTGGTTCGCGCAAGGAAT 766
Qy 474 LeuIleGluAspSerProAspPheArgIleLeuGlnLysLysProArgSerMetAlaGlu 493
Db 767 GAGATCAACAGACTCTCCGGACTTCAGAAATCTCCGAAAAAACCCAGATCGACAGCTGAA 826
Qy 494 ProGluIleLysThrLeuValIleAlaGluValPheAlaGluAspSerGlyCysPhe 513
Db 827 CCCGAGGAGATCTGACCCCTGTCTGAGAGTTTCCCTGAAGATGAGGCATCTT 886
Qy 514 ThrCysThrAlaSerAsnLysTyrGlyThrValSerSerIleAlaGlnLeuHisValArg 533
Db 887 ACCTGCTCAGCCACAATGACTACGGCTCAGTGACACGACCTGACACAGCTTGTTATA--- 943
Qy 534 GlyAsnGluAspLeuSerAsnAsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeu 553
Db 944 -----ACCTCCGCCAACACAGAGAACTGTAGTTACGACTCAACAGGAGAACCCACAGT 997
Qy 554 AlaAlaIleGluProGlnProSerProPro----- 563
Db 998 GATCATTCCAACACTTTCCACTCTCCCTCCGATCTCGGAGCAGGCTCTCTATGAGCTG 1057
Qy 564 ---HisSerGluProSerValGluGlnProProLysProLysLeuGluGlyValLeu 582
Db 1058 GCATCCAGAAACCATCTGAAATCCAGCAGGTGAACAGCCCAATTTA----- 1105
Qy 583 ValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArgValHisPheAsnLeu 602
Db 1106 -----GGATTTAGCATGGCAGCTCTCAAAATGCAATTCAACACT 1144
Qy 603 ProGluAspAspLysGlySerGluAlaSerSerGluAla-----GlyValValThrThr 620
Db 1145 GCAGAGGAGGAGACCAACGGAGTCCATCCCGACCATGAGTCAACGGGCTGATT----- 1198
Qy 621 ArgGlnThrArgProAspSer**GlnGluArgPheAsnGlyGlnAla-----Thr 637
Db 1199 -----AATGCAAGCTTATGCAATAAA 1222
Qy 638 LysThrProGluProSer-----PheProValLysGluProProProValLeuAla 654
Db 1223 TCTCTCCAACACCACTGCGCTCTTACCCACTTAGGAACCAACCGCCACTCTCTGCCC 1282
Qy 655 LysProLysLeuAspSerThrGlnLeuGlnGlnLeuHisAsnGlnValLeuLeuGluGln 674
Db 1283 AAACCAAACTGGATCTCTGAAGCTTCAGCAGCTGCAGAACCAACGAGTGGCGCTGGAACAG 1342
Qy 675 HisGlnLeuGlnAsnProProProProSerSerProLysGluPheProPhe**MetThrVal 694
Db 1343 GAGGCTGCGGCTGCGCCCA-----GCGCCCGCAGGTGTCCCTTGCAACAGCAGCAGC 1396
Qy 695 LeuAsnSerAsnAlaProProAla----- 702
Db 1397 AGCGGCGACGCGCGCCCGCTGCGCCCTTCCCGCCCGCCCGCCCGCCCGCCCGCTTCCGGAG 1456
Qy 703 -----ValThrThrSer***LysGln 709
Db 1457 CTCGGGCGCTGTGCTGCCGTGCCCTCGGAGCCCATGAGCGCTCTCGGCTCCCGCGCT 1516
Qy 710 ValLysAlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePheProSer 729
Db 1517 ACCGCCATGCGTCTCGGCTCTTCAACTACGCGCGCCCTTAAGCAGTTTCATCGCGCGC 1576

Qy 730 ThrAsnThrThrAlaAlaThr-----ValAlaProSerSerSerProValPheThrLeu 747
Db 1577 CAGAACCTGGGTCCCGATCGGGCTGCCACACGCAACTCCAGCCCG-----AGTCC 1630
Qy 748 SerSerThrProGlnThrIleGlnArgThrValSerLysGluSerLeuLeuValSerHis 767
Db 1631 TCCAGCTTGGGTCCCATCTGCCCGACGCGCCAGGCCCTTCGCGCGCGCGCGCGCG 1690
Qy 768 ProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAsnGluProLeuProPro 787
Db 1691 CCCTTCGTGAACCCGAGGCCATGTGGGCGCGCTCTCGCCCTCGCGCCCGCCCGCCA 1750
Qy 788 GlyProThr-----GluPro 792
Db 1751 CCTCGGCTCTCAGCCCTCTGCGAGCTACCCGGTGGCCGACGTGTTCCTCCACTGCCACG 1810
Qy 793 ThrProProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysVal 812
Db 1811 CCGCGCGCACCG-----CTACCTAGCTCCACCTCA-----CACTGGCGC 1849
Qy 813 SerProIleProValSerProThrSerArgIleGlnAsnProValAlaPheLeuSerSer 832
Db 1850 TCGCCCGCGCTTCGCGCCCGCAG-----CAGACGCGCGCGCGCTTCTCTCAGCGCC 1900
Qy 833 ValLeuProSerLeuProAlaIleProProThrAsnAlaMet**LeuProArgSerAla 852
Db 1901 CTGCTGCCCTCTCAGCCACCGCCCTGCTGCTCAACGCCCTGGGGCTGCCCAAG----- 1954
Qy 853 ProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProValAsn 872
Db 1955 --GGAGTCAACCCCGCGGGTTCCCAAGAAAGTCCAGTAGAACTGCTAGAACTGCCTCT 2011
Qy 873 AspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysLysIleThr 892
Db 2012 GATGAGGAGATTCAAGGCACAAAGATGCTGTCTCAAGACCTGGAACGGAAGCTTCCG 2071
Qy 893 PheSerAspValArgProAsn----- 899
Db 2072 TTCAGGAGGAGCACTTCTGAACAATGGCCACCGAGGCTTAACCTATGAGGAAGAATGCT 2131
Qy 899 ----- 899
Db 2132 CGCGCGCTGTGGAGCGCAGCGCAAAACGCTTCAACATCCAGGAGCCAGAGAAACG 2191
Qy 900 -----GlnGlnGluTyrLysSerSerPheGluGlnArgLeuMetAsnGluIleGlu 917
Db 2192 GCAGCCCAATCAGGAGTACAAAGTCTCTAGCTGCGAGCAGAGGCTGATTAGCGAGATTGAG 2251
Qy 918 PheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAspGluIle 937
Db 2252 TACCGGCTGGAGCGCTCCCTGTGGATGAGTCCGGAGACGAGGTGCGAGATCCAGATGTG 2311
Qy 938 ProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThr 957
Db 2312 CCGTGGGAAACGCAACAGCTCCCTTCTTGTAGATGAAGCTGAAACACTACAAGATCTTT 2371
Qy 958 GluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysValThr 977
Db 2372 GAGGGGATGCGCGGTGACTTTTCCAGTGTGAGTGGGAAATCCAAAGCCAAAGATCTAT 2431
Qy 978 TrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGlu 997
Db 2432 TGGTTTAAAGATGGGAACAGCATTTCTCCGAGAGCGCATCTACACCATTCAGAGAGAC 2491
Qy 998 GlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyr 1017
Db 2492 CTGATGGGACCTGCTCTCTCCACACCGCGCTCTACCTCCCTAGACGACGATGGGAACCTAC 2551
Qy 1018 ThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGln 1037
Db 2552 ACCATCATGGTGCACACCCCTCAGGGTGCCTCAGTTGTACAGGAAGGCTATGTGATACAG 2611

```
QY 1038 SerLeuProIleArgSerArg-----LeuThrSerAlaGlyGlnSerHisArgGlyArg 1055
Db 2612 GCTGTCAACCAAGAGCGCGAGTCGCCGCTCCCTCAGGCGCATCTCCATGCCAAGG 2671
QY 1056 SerArgValGlnGluArgAsp-----LysGluProLeuGlnGluArgPhePhe 1071
Db 2672 CCTCGCTCTCGATCACGGCAGTCAGATGAACACGACCCATTTCAGGAGCGATTCTTC 2731
QY 1072 ArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArg 1091
Db 2732 AGACCTCACTTCTCGAGCTCCTCGAGACCTGACCGCTCAGGAAGGCAAGCTCTGCAGG 2791
QY 1092 LeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGln 1111
Db 2792 ATGGACTGCAAGTCAAGTGGTGTACCAACCCAGATCTCAGTGGCAACTAGATGGAAG 2851
QY 1112 ProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerLeu 1131
Db 2852 CCCATACCCCGCCGACAGTCTCACAGATGCTGTGCTGCTGAGATGGGGTCCACTCCCTC 2911
QY 1132 LeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLys 1151
Db 2912 ATTATAGACCATGTCACGTCGCCGAGCGCGCATCTACATGTATTGCCACCAACAGA 2971
QY 1152 ThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysAla 1171
Db 2972 GCAGGACAGAATCGTTTAACTGGAGCTGTGTGCTGTCTGAAGAGCACACAAAGGCC 3031
QY 1172 ProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgLeu 1191
Db 3032 CCTGTGTTATGAGGAAGCTACAGAACACGGGGGTGTGTGATGATGATACCCAGTGGCGTG 3091
QY 1192 GluCysArgValIleGlyMetProProValPheTyrTrpLysLys 1207
Db 3092 GAATCCCGTGTCTCGGAGTGGCGCCACTCAGATATTTTGAAGAAAAA 3139

RESULT 9
AC125500/c
LOCUS
DEFINITION AC125500 136493 bp DNA linear HTG 23-AUG-2002
Takifugu rubripes clone 206P3, WORKING DRAFT SEQUENCE, 2 ordered
pieces,
AC125500
AC125500.2 GI:22450574
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS
SOURCE Takifugu rubripes
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 136493)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Hagnigni,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masello,C., Maskeri,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Paquirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
Stantipop,S., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wecherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 136493)
Green,E.D.
Direct Submission
Submitted (28-JUN-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 136493)
Green,E.D.
Direct Submission
Submitted (23-AUG-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
```

COMMENT

On Aug 23, 2002 this sequence version replaced gi:21622707.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@hgri.nih.gov
----- Project Information
Center project name: dif
Center clone name: 206P03

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 136144 bases at least Q40
Consensus quality: 136285 bases at least Q30
Consensus quality: 136382 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 136393; sum-of-contigs
Quality coverage: 11.03x in Q20 bases; agarose-fp
Quality coverage: 10.51x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 135196: contig of 135196 bp in length
* 135197 135296: gap of unknown length
* 135297 136493: contig of 1197 bp in length.

FEATURES

source
1..136493
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="206P3"

misc_feature
1..135196
/note="assembly_fragment
clone_end:T7
vector_side:left"

misc_feature
1..11443
/note="clone overlaps with GenBank Accession Number
AC124911 clone 209H6 (center project name die)"

misc_feature
135297..136493
/note="assembly_fragment
clone_end:SP6
vector_side:right"

BASE COUNT 38625 a 29019 c 29646 g 39103 t 100 others
ORIGIN

Alignment Scores:

Pred. No.: 3,53e-64 Length: 136493
Score: 1953.50 Matches: 779
Percent Similarity: 24.00% Conservative: 173
Best Local Similarity: 19.64% Mismatches: 304
Query Match: 28.35% Indels: 2716
DB: 2 Gaps: 66


```
US-09-818-990b-2 (1-1320) x AC125500 (1-136493)
QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85826 ATCAGAGAACACACA--GAGCAGCTCAGTCCCTGCTCAGCTCTAAGGGAGAGCTAC 85770
QY 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85769 ATGGCTGAGGCC-----AGGGCCAGCATCGTGCACAGCGCGCAGATGCTCTCTCTCG 85716
QY 41 ProCysHisPheGlySerPro---SerGlyAlaAlaGluGlyGlyGluGlnAspAsp 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85715 CACCTTCAGTTCACCTGCCCTTGAAGGGTAAAGCAATGATGAGCCACGCGCAT 85656
QY 60 Leu-----ProAspLeuSerAlaPheLeuSerGlnGluGluLeuAspGluSerValAsn 77
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85655 GTCCATTTCCAGTCTCTCAGCTCTCCAGCCAGAGAGAGCTGGACAAGATTGAAC 85596
QY 78 LeuAlaArgLeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArg 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85595 CTGGCCCGCAGCCCATCGACACGAG-----CACAGGGCAGAGGTCAGC 85551
QY 98 LysArgLeuSerProAspGlnMetLys-----HisSerProAsnLeuSerPheGluPro 115
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85550 GCCTCTGTACCCCGCAGCAGCCCTGTCCAAACCCCACTCATCTCAGCCTCATCTGAACC 85491
QY 115 ----- 115
Db 85490 CTGCCTGTCCATTGGCTAACCACTCAGTGGCCCCCTTACCGCCCCCCCAACCAAGCCA 85431
QY 116 ----- 116
Db 85430 CTTCTCAATTCAACAACTGCCAGCAGCTCACAATATCTCTTCAACCAAGCAGCAGATG 85371
QY 124 SerProThrSerSerLys----- 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85370 CACAAGACATCTTCAGGGCATGACATGATCAGGAACAGCGGGACAGCGCGAAGGA 85311
QY 130 ----- 130
Db 85310 ATCAATGATCCCAACAGGGGGGAGAACAGCTCCCTATTAACCTGGAGACCCAGTCCAAG 85251
QY 148 LysValPheLeuAsnLysAlaAlaAspPheIleGluLeuSerSerLeuPheLysSer 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85250 AAGAATTCCTCAACAGCGCAGGACTTATCGAGGAGCTCTCTCTCTGTTCAAGGCT 85191
QY 168 HisSerSerLysArgIleArgProArgAlaCysLysAsnHisLysSerLysLeuGluSer 187
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85190 AACAGCTCCAAACGGATACGACCCAGATCTTGCAAAATCCCAACCGGACTAGG-----AGC 85137
QY 188 GlnAsnLysValMetGlnGluAsnSerSerPheSerAspLeuSerGlu---ArgArg 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85136 CAGAACAAG-----AGCCACAATGATGATCATTTGATCCCTGGATGAGGATACCAGG 85083
QY 207 GluArgSerSerValProIle-----ProIleProAlaAspThrArg 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85082 GAGCGTGCCCTCATGTCACAGAGGAGGAGGAGGAGGCGCTTACCCTGCT----- 85032
QY 221 AspAsnGluValAsnHisAlaLeuGluGlnGln----- 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85031 -----GTTGGCCACCAACAGAGGAGGAGCGCGGACGCTGGGATCGGCCACTTTCAG 84981
QY 232 -----GluAlaLysArgArgGluAla-----GluGlnAlaAlaSerGlu 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84980 GACTGCAGGATTACTGAGCGCTGCAGCAAAACTGTCTGACCAAGGACGCTCGAGAAGAG 84921
QY 245 AlaAlaGlyAspThrThrProGlySerSerProSerSerLeuTyrTyrGluGluPro 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84920 ACCGAGGTCGTCTCTGACAGAGAGCTCTACCCAGCA-----GAGCCT 84876
QY 265 LeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluValProGluGlyThr 284
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84875 GTATGTGAGCGCGCTCATTTTATCCAGAAACTGAAGAGCAGGAGATCGCCGAGGCGAGC 84816
QY 285 ArgValGlnLeuAspCysIleValValGlyIleProProGlnVal----- 300
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84815 AAGTGCAGTTGGACTGCATCGTAAGAGGATTCCTCGCTGCCTGAAGTCAGGTGAAGCAAC 84756
QY 300 ----- 300
Db 84755 CCAACAGACACCAGGCTGACCTGTGACTTCTTAATCATGTTCTCTCAAAACGAGCCGAAC 84696
QY 300 ----- 300
Db 84695 CTCGAGGATGTTTGTCTCCACGAGTACGCTGACGATCTACAGAGTGGAGCTGTGGGTCA 84636
QY 300 ----- 300
Db 84635 CGTTTCATTCAAACTCACTGAATCTGAAGTAAATATTTCATATATTAAACCATTTATATTA 84576
QY 300 ----- 300
Db 84575 TATCATCCACTCTAACATTCGGTTTAGTTCGGTATTAGTTGTTTAGTAGAGTTTTCAA 84516
QY 300 ----- 300
Db 84515 GAAGTTGTTGGTCAAGCTTTCATCGGCCAGACTTGTAGTAAATGTATGAAGTGTGCC 84456
QY 300 ----- 300
Db 84455 ATGAGATAAACCTAAAGTGAACATTGTGAAATACAGCTGTTGTTTCAATAGTATTTC 84396
QY 300 ----- 300
Db 84395 ACACAACTGAAAAGATGATTAAATCAGAGGCCAAGAAATTGAAAACCTGGAGAACATCAC 84336
QY 300 ----- 300
Db 84335 ATGAGCAGAGCTGCGATGATGCTGAAGTTCCTCAGTCATGATGCTGTTTTCGATACATT 84276
QY 300 ----- 300
Db 84275 GCAGACAAAGTGACATGTGATGACAGCTGTTCTGAAACGGTGCATGCTGATTAACAG 84216
QY 300 ----- 300
Db 84215 AACCCCGAGGTTCCCTGCTTCATCAGATGATGCTCCACAAAACAGCCACACATGT 84156
QY 300 ----- 300
Db 84155 TGACCCCTCGGCAGAAATGAAACATTTTTCAGCAAGTGAGATGGATTTAGCCATCGTT 84096
QY 300 ----- 300
Db 84095 CTGGCTTATTCGTGACACTTATGTTAACTTGTTAGTCACAGGGTGTACAGACTATA 84036
QY 300 ----- 300
Db 84035 GGGGCGATGGCTGTGGTCTCAGTAACATTTGTAGACCATGACCTGACACTTGTCTCG 83976
QY 300 ----- 300
Db 83975 AGAAATGTCGCTCTCTGCCATTTTAGCCCGGACAGCAGAGCTGTAGTTGTGTC 83916
QY 300 ----- 300
Db 83915 TGATAATGGAACCGAGGAGCTCTTGACAGCAGCGTGACGTGTGTAAGGTGCTTTTAC 83856
QY 300 ----- 300
Db 83855 ACTGACATTCCTCAGCCAGTATCTTTATTTTCTCATTTTAGAGAGCTGCTTTGTTCTTTT 83796
QY 300 ----- 300
Db 83795 TCTGCATGTCCCCAAATGTGCCACTAGTTTGTGGTGGGTTTGCACCTGTGCCACGA 83736
```

QY 300 ----- 300
Db 83735 ATGTTTATTTATGATGTTTAAACCTAAACACTGCCGCTGTCATCACTAGTTGTATTTTA 83676
QY 300 ----- 300
Db 83675 AAGCCAAAGTTTCAGGCTCAGATTTCCCTGTCCTGGTGTGTGTCAGAGTGAACACTGCGTG 83616
QY 300 ----- 300
Db 83615 TATTACAAGTCTATGTTTAGCACTTTGGTATGAGATTGTGTCACTAGTCAGTCACTG 83556
QY 300 ----- 300
Db 83555 GGCTCCAGTGGGAATCTCCTCGGTGGCTCTCACATCCGTAGGAAATAACCCACAGGT 83496
QY 300 ----- 300
Db 83495 GCAGCAAGCAGGGTCTCAGCCTCACAGACCGTCCAGCTCATCTACTCATGCTTTCAA 83436
QY 300 ----- 300
Db 83435 TTTAGACCTTACTCATTTTGAATGATATAGCTCACCTCAGATCTGGCTGACTAGAGAA 83376
QY 300 ----- 300
Db 83375 TTAAGCTGAAGCTACGCTGATTACTGGCTTGTGTTTAACTCCTGGATTGTTGCTT 83316
QY 300 ----- 300
Db 83315 GTAGAATATCAGACAACATAACTCTGAGACAACATAACTCTGACCTTCCCGTACGAATGT 83256
QY 300 ----- 300
Db 83255 TCTACTGCCCTGATTTAGAAATCCATCTGCACATTCCTCTGTGTTTCAATGGCCTGATCT 83196
QY 300 ----- 300
Db 83195 GGTAAGCATCAAGTTTCATTTGTTGAATAGATTGAGCATAGTCTGGTCAGGATG 83136
QY 300 ----- 300
Db 83135 TGGAAAGCTGGACGCTTTATTTATTAGCATATGAGAAATGAACGGGCTGTTTACGACTGA 83076
QY 300 ----- 300
Db 83075 GCTCAGAGCTTTGGGAATTTACTGCTATTTATTTTGTGTTTGTGTTTCCAAAAGTATGG 83016
QY 300 ----- 300
Db 83015 AGCCTTTTATGTCATATATAGCCAGNATGATTAAAGTATAGAGATGATTTGGCTTCAG 82956
QY 300 ----- 300
Db 82955 ATGTTTGTGATGTGTTGCTTCTTCATCATAGTCTTTAAATGACTGGGATCTTTTCCCCAG 82896
QY 300 ----- 300
Db 82895 TATTAGTGTAGTAGAGACCAGTCTCAGGGGGCCAGTTTAAGAGAGAGACGTCCTGCTGA 82836
QY 300 ----- 300
Db 82835 AACAAATTGCAGTCCGGACATTTATTGGAATGGAATAGCCAGACCITTTGTGATTATTA 82776
QY 301 ----- ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGI 319
Db 82775 TTGCAGATGTTCTGTGAGGGCAAGAGCTGGAGAACAGCCCTGATATTCAGATCATCAA 82716
QY 319 nAlaGlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTy 339
Db 82715 CGACGGGAGCTCCACTCTCTCATCATCTCAGAAGCGTTTGAAGAAGACACCGGGCGCTA 82656

QY 339 rSerCysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGI 359
Db 82655 CTCCTGCTTTGGCCTCTAACATCTACGGCACCGACTCCACGCTGAGCTGAGATCTACGTGA 82596
QY 359 u----- 359
Db 82595 AGGTGGTGTCCGTCCTCTCCTCACGGCCACAGCTGTCTTTACCCTGTCTGCTGATGAAGA 82536
QY 360 ----- GlyValSerSerAspSerGlu----- 367
Db 82535 ATGTGCTTTTATTTGTGCAGGTGCCCTCCTCTCTGACTCCGAGGAGAGCATCATTTTGA 82476
QY 367 ----- 367
Db 82475 ACACACAGCCAGTGTGTTTAAAGCCCATTTTCAATTTAACTATTATGAGCCACTTTATT 82416
QY 367 ----- 367
Db 82415 TGATCGCGAGTATGCTCTTTCTGCTATTTGAAGTGTTCAGCCAAAATCTTTACCCC 82356
QY 368 ----- Gly-AspProAsnLysGluGluMetAsnA 377
Db 82355 CATTTTCAAGTCAGGCTCTCTCTGCAGAGCAGGAAGACCTCTGTATCTCTCTGAAGCAG 82296
QY 377 rgIleGlnLysProAsnGluValSerSerProProThrThrSer----- 391
Db 82295 CCACTGAGGAACACAGCAGACGCCGCCCCACCACTAACCAAAATACCCCCAGAAC 82236
QY 391 ----- 391
Db 82235 CTCTGCAAAAATACTTCCAGTCTCAGATTTCTCCAAACCCCGGGAAGCCTGCAACGC 82176
QY 391 ----- 391
Db 82175 AGGTGACCGAGCAGCGGATCCAAACGCTATCTGTGAGGGGATTTCTGCTTTGTACAGAA 82116
QY 392 ----- AlaValIleProPro----- 396
Db 82115 GTGACCCAGAACATCTCTGAAAGTCAGACAGCCATTTTCCACCGTGTCTCTTGTCTGCCG 82056
QY 397 ----- AlaValPro----- 399
Db 82055 CGCAGTCCGCTCAGCAGCCACCTCGCTGCTTTTGAAGCGAGTTCCATCCCTCCACCTG 81996
QY 400 ----- GlnAlaGlnHisLeuValAlaGlnPro----- 408
Db 81995 CTTTGACACCTCCACCAGGCATCCACCCCTGAGGTACAAACCACCATCAACTTCCGCC 81936
QY 409 ----- ArgValAlaThrIleGlnGlnCysGlns 418
Db 81935 ATATTATTAGATTAGCTCTGTTGTTTAAACCACCTTATCAACTAATATGCCAGATCAG 81876
QY 418 erProThr----- AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValP 437
Db 81875 CCCCAGCTGTAATTTATTCGAAGTCTGAATGGCAGCCTGTCTATGCGCAGCACCATAT 81816
QY 437 heThrLysMet----- 440
Db 81815 TCACAAAGGTACGTTCTCTGCGTGTCTCCACTCTACTGTTCTATCTGCTTCCATCTA 81756
QY 441 ----- LeuGlnAsnLeuSerAlaSerGluGlyGlnL 451
Db 81755 CCATCGACTGTTGTTGTTGCTCCCTCCAGAACCTGCAAGAGCTCTATGCTCCGAAGCCAA 81696
QY 451 euValValPheGluCysArgValLysGlyAlaProSerProLysValGluTrpTyrArg 471
Db 81695 TGTGTGCTAGAGTGCCTGTGAAAGAGTACCGTCCCGCAGGTGCGCTGGTATAGGA 81636
QY 471 luGlyThrIleLeuGluAspSerProAspPheArgIleLeuGlnLys----- 486
Db 81635 ATGAAACATGATTGAGGATTTCCCCAGATTTTCAGGATCTTACAGAAAAGTAAGGCATAAT 81576
QY 486 ----- 486

QY 658 ----- 658
Db 79419 ATTAGTCTGAGACATTTTACATGAATTTGATAAATAAGTAGTGATGCAATTCAAATTTGAAA 79360
QY 658 ----- 658
Db 79359 GCTGTAGTGTAGTAATGACAGGGTCTTTGTTGTATCCATAAAGGAGCAGCATATTCCT 79300
QY 658 ----- 658
Db 79299 CACTCTGGGGTTTACATGGATGGTTTATTAGAGGAGTGCAGAAAGCTTCCCCCCTCAC 79240
QY 658 ----- 658
Db 79239 CTTTGACCATGTATAGTGTGATTTTAAAGCTGGGCCATTTGTTGTTTAAATTGAC 79180
QY 659 --- AspSerThrGlnLeuGlnGlnLeuHisnGlnValLeuLeuGlnGlnHisnLeu 677
Db 79179 AGGGACCCAGCTCAGCTCCAGCTTCTGCACAACCCAGGTCTGTTGGAGCAGCAGGAG 79120
QY 678 GlnAsnPro----- 680
Db 79119 ACGNACCACAGACACAAACTCAACTCAACTCAACTCAACTCCACCCACACGACGTCCAG 79060
QY 680 ----- 680
Db 79059 GTCCAGCCCGAGGTCAGACCTCCGACAGGTCTATCTGCTCTTCTCAAAATAAACCGG 79000
QY 681 ----- ProProSerSer----- ProLysGlu 687
Db 78999 GGTCCTCAATCTCCACATTTCCACACCCGACACCTCCCTCTACAGCTCTCCCGACCTCT 78940
QY 688 PhePro----- Phe***MetThrValLeuAsnSerAsnAla 699
Db 78939 GCCCCTCTCGCTCCNATGGTCAACACCACCTTTTCTCCCCCTCTCCAAACCTCTCCTGCT 78880
QY 700 ProProAlaVal----- 703
Db 78879 CTTCTCAAGCTTGTCATATTCCTGCTCTCCTCAGCTAAAAAAGCTCCTACTGGTGACATCC 78820
QY 703 ----- 703
Db 78819 CTTCCACCTCCCCACAGATAGGAGAGCTGCTCCTCAGACATCCCCCCCCGTAAGTCAG 78760
QY 703 ----- 703
Db 78759 ATGGATACCTCCACTGCCCCACCTCTGCTGTGTCCTCCCTTGCNAGAGCTGGACCCACA 78700
QY 704 ----- ThrThr 705
Db 78699 GCAAACTCTCTGACTCCCTGGCACCAAGCTGAGTGCAGCGCCGTCGATCCACCCCA 78640
QY 706 Ser***LysGlnValLysAlaProSerSerGln----- 716
Db 78639 TCCCTGCAGCAGATCAGCTCTCCAGCAGCCAGCTGAGGACACCCATGCTTCTCTCATG 78580
QY 717 ----- ThrPheSerLeuAlaArgProLysTyrPhePheProSer 729
Db 78579 ANTCTCACACCTCTCCTCCCTTCAACTATGCCCCGTCCAAAGAATTCATA----- 78526
QY 730 ThrAsnThrThrAlaAlaThrValAlaProSerSerSerProValPheThrLeuSerSer 749
Db 78525 ----- GCTGCACAGACTTCTCCCCAGTCAGGAGCCCCCTCTTAACAGAAATCCCA 78475
QY 750 ThrProGlnThrIleGlnArgThrValSerLysGlnSerLeuLeuValSerHisProSer 769
Db 78474 GTACCTCTGCTGCGAGGAATCTGCTGCAGAGCTCAACTCTCAGAGGCAAGTTCTCCAC 78415
QY 770 ValGlnThrLysSerProGlyLeuSerIleGlnAsnGluProLeuProPro----- 787
Db 78414 ATTCACCTTTTCTCCCA-----CCCCCGAGGGTT 78385

QY 788 ----- GlyProThrGluPro----- ThrPro 794
Db 78384 TTTCCACACGGTTCTGATGCTCTCCACACGCTCCATCCTCCTGTGCTCTTCGCC 78325
QY 795 ProProPheThrPheSerIleProSerGlyAsnGlnPheGlnProArgCysValSerPro 814
Db 78324 ACACCGGATCGTCTCTCTCTCTGAGC---AGCTCGTTTCGCCATCCGAGCTCAGTCGCC 78268
QY 815 IleProVal--- SerProThrSerArg----- IleGlnAsnProVal 827
Db 78267 CTACATGCATCCTCCCCACCTCCAGAGCTCCACTCCGAGCCCTATTTCAGATCCCGTG 78208
QY 828 AlaPheLeuSerSerValLeuProSerLeuProAlaIleProProThrAsnAlaMet*** 847
Db 78207 GCAITTTCAAGTTCGCTCCCATCCTCTCCCTGGCTCAGCCCAACCACTCGATGGC 78148
QY 848 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAla----- 861
Db 78147 CTGCCGAAGGGGCTCCGTC-----GGGTACGCTTTTCCCAATGACCTGTCA 78100
QY 862 ----- LysLys 863
Db 78099 TTCGTGACCGATGGTGCAGCCGATTCAGCTCTTTCATGCTTTCAGGCTTAAAAAG 78040
QY 864 AsnThrLysSerProGlnProValAsnAspAsnIleArgGluThrLys----- 880
Db 78039 AGCTGCCAAAGGACGACCTCAATCAACAGATAATGTTCTGTAAGCAAGCAAGAGATTCTC 77980
QY 880 ----- 880
Db 77979 CATGAATTTGAAAAAGCTTCGATTTAGAGAGGAAACTCCACATTTTGCACATCAACAG 77920
QY 880 ----- 880
Db 77919 GTACTATATACTCTCATGATTTATTAGCTGATTTCCCTCCATCGGTAGCAGTGGTAAAT 77860
QY 880 ----- 880
Db 77859 GGAATGATTTCAAAGACAAACAAATACCTAAGACAAACAAATAGCAGCTGAAGTC 77800
QY 880 ----- 880
Db 77799 CTCCTTTATCGAAATGCTGTGCGATGATTTCAATTTTCAATTCATCCCTGTTTAGATATT 77740
QY 880 ----- 880
Db 77739 TATGTCTAATGCTCGGCATTGGATGTTGGTAGATTTGAATTTGACACAAGGTGAAG 77680
QY 880 ----- 880
Db 77679 TATGTGAATGCTAGCTGCCATCTAACCAATGCTGGTATCTGTTTAACTCCTTTTAAAGG 77620
QY 880 ----- 880
Db 77619 ATAATTTGGTTGATCTTAATAATGTTATTAAGGAAGCTGCCCTCACAATGTCTGA 77560
QY 880 ----- 880
Db 77559 GAAATGAGGAAGAACCCATGTTCAAGCCGACGTTCTCTGACTTCATCTTGTACTGTCAAT 77500
QY 880 ----- 880
Db 77499 TACCCGTGTGTACCATCACCTAACGATGCTAGTAGAGCCCACTTGTGTTCTTCCA 77440
QY 880 ----- 880
Db 77439 TCAGAAATGTAATGTGTATGTAGTGTTCATCCTCATCCCTGCTCTCATAGAAG 77380
QY 880 ----- 880
Db 77379 CTGAGTGTGGGGGAAACAGCAGCAGACCTGTCGATCAAACTTTCCTGCTACTGTCT 77320
QY 880 ----- 880

Db 77319 GTTAGCTATGACGAGGTACAAGGATCCTCCACAGCGCGTGTCTCGCGTGTACGGGT 77260
Qy 881 -----Asn-AlaValIleArgAsp----- 886
||| |||||::: :::
Db 77259 CATTAATAATATGCACTAAAGAGAGAACCGCAGTTCTCTTAACATTATCTATAAAGCT 77200
Qy 886 ----- 886
Db 77199 TTATCTAATATTACACATAACATTTACTAATAATTTCAGGGAATTCAACTGATTTTAGGGTA 77140
Qy 886 ----- 886
Db 77139 AGAATCCATTAAACTTATTGAATGTTGCTGCTTCTATACAATTAACCTCATCGAAACGTC 77080
Qy 886 ----- 886
Db 77079 TTGTTTCCTTTTTGTGATTTTTTTTTCACATAGAAAATACCCTTATTTCCCAACCACT 77020
Qy 886 ----- 886
Db 77019 GTTATTAAATTTGCTTTTGTCTGCTCTGAGGTGATGTAAGCAATTTCTTAGTGTGACTC 76960
Qy 886 ----- 886
Db 76959 AAACAAGGAAGTGAGTGTGGATGTTTTTCTTTAAATTTTGAGTGTCTCAGTTTCATCATC 76900
Qy 886 ----- 886
Db 76899 TCTGTTCTCTGTTTCTTCGCCAGTGTCTGCCAGCATGTCTGCTGCATGTCTATGTGAC 76840
Qy 886 ----- 886
Db 76839 ATGCCAGTTTGACAGTCTGCTTATAAATTCCTACTCTAATGCTACAGTTACTAGATGAC 76780
Qy 886 ----- 886
Db 76779 TGCACATAATCAGACTTGTAGTAAGAGCTTGTGCTGCATCTTTGATTACGTTTGTATGGAC 76720
Qy 886 ----- 886
Db 76719 AGTTGCCAAAATCCAAAACACATCTCCTTAFTTCCCTGTTTAAAGTCTTTATATAAATTTATCA 76660
Qy 886 ----- 886
Db 76659 TTCTGATATAATTTGGAGATCTTTGGAGAGCAFAAAGTGGNAAGATTGTAAAATTTCACT 76600
Qy 886 ----- 886
Db 76599 ATAACAGAAGAAGAGGCTTTGGTGAAGTAAAATAAATAATATTGTTGGCAACCAACCA 76540
Qy 886 ----- 886
Db 76539 TAAGAGTTGCTATCATCCATCTGCAAAATCCTGTTGTAAAGTCTTTATATAAATTTATCA 76480
Qy 886 ----- 886
Db 76479 GTTTTTCAATGTTGCCAAGCTAATTTGGACCACAGACAGACAATCATCTAACACCATCA 76420
Qy 887 -----LeuGlyLysIle 891
||| |||||::: :::
Db 76419 CCAGCCTAGGAGTTATAGGAGGCTTTAGGAGGTACCCGACACTAACTAGGACGTAATAAT 76360
Qy 891 eThrPhe-----SerAspValArgProAs 899
||| |||||::: :::
Db 76359 CACCTTTAAAGCTAACTGACAGTTAGACGGTGTGATACAGGGCACAGAGGAAGACACAGA 76300
Qy 899 n----- 899
Db 76299 GGACTGAACAGAGGAGTCACTTCAATTTTTTAAGAGAAAGCAAAATATTTGGCCTATACCT 76240
Qy 899 ----- 899

Db 76239 TAAATGATCAAAATCCCAACAGATTAATCCTGTTGGAAACAGGAAGATGCATTTGTAT 76180
Qy 899 ----- 899
Db 76179 TTTAACTCTGAGCATCAGCTTCTAAGATACCGTATGTCATGTTACTGTCACGACACAGTTCA 76120
Qy 900 -----GlnGlnI 902
||| |||||::: :::
Db 76119 GTGCTGGATGCATCATTTGGCTTTGACTTTACATCATCGGTTTCTTTTTCCTCCAGGA 76060
Qy 902 uTyrLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluAr 922
||| |||||::: :::
Db 76059 GTCAAAAGTCTCCAGTTTTCAGCAGAGGCTCATCAGTGAGATAGATTTCTGTTAGAAGC 76000
Qy 922 gThrProValAspGluSerAspGluIleGlnHisAspGluIleProThrGlyLysCy 942
||| |||||::: :::
Db 75999 AACGCCGTGGAGGAATCGGATGATGAAGTTCAACATGATGATGCCCACTGGGAATG 75940
Qy 942 sIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProVa 962
||| |||||::: :::
Db 75939 TATAGCCCCATCTTTTGAGAAGAGCTGAAGAAGCTTCAAGCTACGGAGGGCGTCCCTGT 75880
Qy 962 lThrPheThrCysLysIleValGlyIleProValProLysVal----- 976
||| |||||::: :::
Db 75879 CACCTTCTCATGTAATAATATTGGAATCCCTGTTCCAAGGT -GAGGAGACACAACTGT 75821
Qy 977 -----TyrTTPheLy 980
||| |||||::: :::
Db 75820 AACCCCCCTTAGGAACAATACATTTATCTGTAAATCCTCAACAGGTTTACTGGTTCAA 75761
Qy 980 sAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgGluGlyAspGI 1000
||| |||||::: :::
Db 75760 AGATGGCAAGCAGATCTTAAAAAGAGCTCCATTACAGGAAGATAGAGAGGGGATGG 75701
Qy 1000 yThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIleMe 1020
||| |||||::: :::
Db 75700 AACCTGTCTTTACATATTGATGTCACGACCAGTCAGTATGATGGAACTACACCGTAAT 75641
Qy 1020 tAlaAlaAsnProGln----- 1025
||| |||||::: :::
Db 75640 GCGCGCTAATCCACA -GGTAGTACACTTCTGTCTGAGTTAGCATGAGCTTCCCTCTGAA 75582
Qy 1026 -----GlyArgIleSerCysSerGlyHisLe 1034
||| |||||::: :::
Db 75581 AGCAAAATCCAAACTCTTGTTAATCACCATCAGGAGCGATCAGTCTCAGGTCATTT 75522
Qy 1034 uMetValGlnSerLeuProIleArgSerArgLeuThr-----Ser-AlaGlyLns 1051
||| |||||::: :::
Db 75521 GATAGTCCAAACGGGACCTCTCAGAACCCGACTCTCCCATTTTTCATTTCTCAGAGGTGAGC 75462
Qy 1051 er----- 1051
Db 75461 AGAAGTCATTTGCAAGTGGAGAACTGGATCATATTTTGACATCATATATCATGTCCTCT 75402
Qy 1052 -----HisArgIlyArgSerArgValGlnGluArgAspLysGluProLeuG 1067
||| |||||::: :::
Db 75401 GCATTACACATCTTTTCAGGGGGCAATCCTCGTACGGGATGCTGAGGGTGAGCAACCC 75342
Qy 1067 lnGluArgPhePheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluG 1087
||| |||||::: :::
Db 75341 AGGAGCGCTTTTTCAGCCTCATTTCTTCAAGCTCCAGGAGACATGCTGGCTCATGAGG 75282
Qy 1087 lyArgLeuCysArgLeuAspCysLys----- 1095
||| |||||::: :::
Db 75281 GAAAACTGTGCGAGCTAGACTGTAAAGTGAAGATCAGGCCAGGGGACCACAGAAATGTG 75222
Qy 1096 -----V 1096
Db 75221 GGAATGGTTTTCAATTATGTTTTTCCCTCTCATTTGATTTCACTGTTTCTTCTCTCCAGG 75162
Qy 1096 alSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProA 1116
||| |||||::: :::
Db 75161 TGAGTGGTCTACCCATCCACAGTTGATGTGTTGGTCAACGGGAACCCATCTACTCAG 75102

Db 52 GGTCCGGAGCCCGCTGCAGCTCCCG-----CTCGTCCGGAGCCGGGAATCGGGC 102
QY 655 -----LysProLysLeuAspSerThrGlnLeuGlnLeuHis 667
Db 103 AGCAGCGGAGGCGCGCCGAGAGCCGAGG---GACCCTCTGAAGCTCCAGCACTCCAG 159
QY 668 AsnGlnValLeuLeuGlnHisGln---LeuGlnAsnProProPro-----Ser 683
Db 160 AACCAATCCGACTGGAGCAGGAGCGCGCTCGGCAGCCTCCGCCAGCCCGCGCAGC 219
QY 684 SerProLysGluPheProPhe----- 690
Db 220 GCAGCGCCCTCGCCGCCCTTCCCGCGCGCCGCCCTTCCCGAGCTCGCGGCTGCAGC 279
QY 691 -----MetThrValLeuAsnSerAsnAlaProProAlaValThr 704
Db 280 CCGCGCGCTCCCGGAGCCCATGAGCGCGCTGCCCTCCCGCTCCGCGCCCGCATG--- 336
QY 705 ThrSer***LysGlnValLysAlaProSerSerGlnThrPheSerLeuAlaArgProLys 724
Db 337 -----CAGTCTCTCGGCTCTCAACTACGCGCGCCCAAG 372
QY 725 TyrPhePheProSerThrAsnThrThrAlaAlaThr-----ValAlaProSerSerSer 742
Db 373 CAGTTTCATCCCGCGCAGAACCTTCGGCGCGCTCGCGCCACGCGCGGCTCCAGC 432
QY 743 ProValPheThrLeuSerSerThrProGlnThrIleGln-ArgThrValSerLysGluSe 762
Db 433 CCC-----AGCTCGTCCAGCTCCCTCGCCCGCTCGCGCCACGCGCGGCTCGGC 486
QY 762 rLeuLeuValSerHisProSerValGlnThrLysSerProGlyGlyLeuSerIleGlnAs 782
Db 487 CGCGCGCCCGCTCGCGCTTCG-----CGCAGCGCTTCGCGCTGAGCGCGAGGCC 537
QY 782 nGluProLeuProProGlyProThrGluProThrProProPhePheThrPheSerIlePr 802
Db 538 CCGTGGGCTCTCTCTCG-CCGTGGCGCGCGCGCGCGCGCGCGCGCTCTTCAGC----- 591
QY 802 oSerGlyAsnGlnPheGlnProArgCysValSerProIlePro----- 816
Db 592 -----CCAGGCTGCTTCCCGGCTCGCGCGCGCGCGCTTCCCGCTGCC 632
QY 817 -----ValSerProThrSe 821
Db 633 CGCGCCACCGCGCTCCCGAGCGCGGACAGCGCTCCCACTGCTCGTCCGCTGCCAC 692
QY 821 rArgIle-----GlnAsnProValAlaPheLeuSerSerValLeuProSerLeuPr 838
Db 693 CGCTTCGGCCACGGCCAGACGCGCGCGCGCTTCTCAGCGCTGCTGCGCTCGCAGCC 752
QY 838 oAlaIleProProThrAsnAlaMet***LeuProArgSerAlaProSerMetProSerG1 858
Db 753 GCGCGCGCGCGCTCAACCGCTTGGGCGTGGCCCAAG-----GGTGTACCCCGCCG 803
QY 858 nGlyLeuAlaLysAsnThrLysSerProGlnProValAsnAspAsnIleArgG1 878
Db 804 AGGATTCCAAAGAGGCCAGTAGAAGTCTAGAAATAGCTCCGCTGAGGAAATTCAGG 863
QY 878 uThrLysAsnAlaValIleArgPLeuGlyLysLysLysLysLysLysLysLysLysLys 898
Db 864 CACAAAGGATGCTGTTTATCAAGACCTTGAACGAAACTTCGCTTCAAGGAGGACCTCT 923
QY 898 oAsp----- 899
Db 924 GAACAATGGCCGCGAGGTTAATACATACGAAGAAGAAATGGCTCGTGCAGCTGTAGTGC 983
QY 900 -----GlnGlnGluTy 903
Db 984 TGACAGTGCAACTGCTCTTTAATATTCAGGAGCCAGAGAGAGAAACACTAATCAGGAATA 1043
QY 903 rLysIleSerSerPheGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgTh 923
Db 1044 CAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATAATAGGTATACAGGCTGAGAAAGTCT 1103

QY 923 rProValAspGluSerAspAspGluIleGlnHisAspGluIleProThrGlyLysCysI1 943
Db 1104 TCTGTGGATGAATCAGGTGATGAAGTTCAAGTTGAGAGATGTGCTGTGGAATAATGAAT 1163
QY 943 eAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGlySerProValTh 963
Db 1164 GGCACCATTTTTCAGATGAAGCTGAAACATTTACAAGATCTTTGAGGGAATGCCAGTAAC 1223
QY 963 rPheThrCysLysIleValGlyIleProValProLysValTyrTrpPheLysaspGlyLy 983
Db 1224 TTTTCATGTAGATGGCTGGAATCCAAAGCCAAAGATCTATTGGTTTAAAGATGGAA 1283
QY 983 sGlnIleSerLysArgAsnGluHisCysLysMetArgGluGluGluGluGluGluGlu 1003
Db 1284 GCAGATCTCTCCAAAGAGTGTACTACCATTCAAAGAGATCTCGATGGGACCTGCTC 1343
QY 1003 rLeuHisIleGluSerThrThrSerAspAspAspGlyAsnTyrThrIleMetAlaLaas 1023
Db 1344 CTTCCATACACAGCCCTCCACCTAGATGATGATGGAATTTATACAATTTATGGTGCAAA 1403
QY 1023 nProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLeuProIleArgSe 1043
Db 1404 CCCTCAGGGCGCATCAGTTGTACTGGACGGCTAATGGTACAGGCTGTCAACCAAGAGG 1463
QY 1043 rArgLeuThrSerAlaGlyGlnSerHis-----ArgGlyArgSerAr 1057
Db 1464 TCGAAGTCCCGGCTCTCCCTCAGGCCATCTCATGTGAGAGGCTCGTCTAGATCAAG 1523
QY 1057 gValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuG1 1077
Db 1524 GGACAGTGGAGAGAAAATGAACCAATTCAGGAGCGATTCTTCAGACCTCACITCTTGCA 1583
QY 1077 nAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValse 1097
Db 1584 GGCTCTCTGGAGATCTGACTGTTCAAGAAGAAACTCTGCAGAATGACTGCAAAAGTCAG 1643
QY 1097 rGlyLeuProProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAl 1117
Db 1644 TGGTTACCAACCCAGATCTAAGCTGGCACTAGTAGGAAGCCCTGACGCCCTGACAG 1703
QY 1117 aSerHisLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuTh 1137
Db 1704 TGCTCAACAAGATGCTGCTGCTGAGAACGGGTGCTCACTCTCTGATCATAGAGCCAGTCAC 1763
QY 1137 rGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPh 1157
Db 1764 GTCAGTGTATGCCGCTATCTACACATGTATAGCTACCAACCGAGCAGCAGACACTCAT 1823
QY 1157 eSerLeuGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLy 1177
Db 1824 CAGCTGGAGCTGTGTTGCTGCTAAAGAGACACAAACCCCTGTGTTTATTAGAA 1883
QY 1177 sLeuGlnAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleG1 1197
Db 1884 GCTCCAAAACACAGGAGTGTGTGATGGTACCCAGTCCGCGCTGGAATGTCTGTATTGGG 1943
QY 1197 yMetProProProValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgG1 1217
Db 1944 AGTGCCACCCCTCAGATATTTTGAAGAAGAAAGTGAATCACTCACTCACAGCACTGA 2003
QY 1217 uArgIleSerMetHisGlnAspThrThrGlyValHisSerLeuLeuIleGlnProAlaLy 1237
Db 2004 CCGAGTGCAGTGCACCCAGGACACACCGCTACATCTGCTGCTCATTCAGGGAGGCCAC 2063
QY 1237 sLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIleValSerCy 1257
Db 2064 AAAAAGATGCTGGTGGTATCTGTGTAGCCCAAGAAATGAACGAGGATTTGTGCTGT 2123
QY 1257 sThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis-----GlnIleProProProme 1275
Db 2124 TACTGCCAGGCTGACGCTTTACACCCAGTGGCATCAGCAGTGCAGCAGCAGCAGCAGCAG 2183

```
Qy 1275 tSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePh 1295
|||||
Db 2184 AAAAGTACGGCCCTCGACCTATGCGACTTTCGACCAGGACTAGACATCAA 2243
|||||

Qy 1295 eSerAlaPheSerMetGluSerThrMetValTyrSerCysSerSerArgSerValVa 1315
:::|||||
Db 2244 AGCAGCGTTCACCT---GAGGCCAACCCATCTCACCTGACACTGAATACTGCGCTTGGT 2300
|||||

Qy 1315 lGluSerAspGluLeu 1320
|||||
Db 2301 AGAAGTGAGGACCTG 2316
|||||

RESULT 11
AC024258 160658 bp DNA linear PRI 14-MAY-2002
LOCUS Homo sapiens chromosome 10 clone RP11-297N15, complete sequence.
DEFINITION AC024258
ACCESSION AC024258
VERSION AC024258.9 GI:20564414
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 160658)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160658)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 160658)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 160658)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On May 14, 2002 this sequence version replaced gi:16905139.
FEATURES
source
1. 160658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-297N15"
/clone_lib="RPC1-11"
BASE COUNT 50744 a 31675 c 30448 g 47791 t
ORIGIN
Alignment Scores:
Pred. No.: 6.9e-49 Length: 160658
Score: 1553.00 Matches: 302
Percent Similarity: 99.34% Conservative: 0
Best Local Similarity: 99.34% Mismatches: 2
Query Match: 22.54% Indels: 0
DB: 9 Gaps: 0

US-09-818-990B-2 (1-1320) x AC024258 (1-160658)

Qy 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
|||||
Db 110106 ATGCAAGACGACATAGAGCTTCTACTTCCATATCTCAGCTTCAAGAGAGAGCTAT 110165
|||||

Qy 21 LeuAlaGluThrArgHisArgGlyAsnAsnGlnGluSerArgAlaGluProSerSerAsn 40
|||||
Db 110166 TTACTGNAACCAACATCGGGGAACAAATGAGAGAGTCGAGCGGAGCCCTCCTCCAAC 110225
|||||

Qy 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspLeu 60
|||||
```

```
Db 110226 CCGTGCATTTCGGCAGCTCCTTCTGGGCGCGCTGAAGAGCGGAGGCCAAGATGACCTT 110285
|||||
Qy 61 ProAspLeuSerAlaPheLeuSerGlnGluGluLeuAspGluSerValAsnLeuAlaArg 80
|||||
Db 110286 CCAGATCTTTTCAGCCTTTCGAGCCAAAGAATATTAGCAAAAGTGTCAAATTTGGCAGA 110345
|||||
Qy 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
|||||
Db 110346 CTGGCCATCAATTAGCACCCTTTGGAGAGCGAGATGAACTCAAGCTAGAAAAAGACTT 110405
|||||
Qy 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
|||||
Db 110406 TCTCCTGATCAGATGAACACTCACCTTAATTTAAGTTTTCAGCCCTAATCTCTGCCAGGAT 110465
|||||
Qy 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
|||||
Db 110466 AACCTCGAAGTCCCAACGAGCTCTAAAGAAAGCCCCAGAGGCAAAAGGCCACAGTAT 110525
|||||
Qy 141 CysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
|||||
Db 110526 TGTCTCAAAACCCAGTCCAAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGAAGAG 110585
|||||
Qy 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
|||||
Db 110586 CTATCCTCCCTTTTCAAAATCCACAGCTCCAAAAAGGATTAGACCTCGCTGCTGCAAAAC 110645
|||||
Qy 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
|||||
Db 110646 CACAAGAGTAAACTGGGAATCTCAAAACAAAGATTATGCAGGAACACAGCTCCAGTTTCTCA 110705
|||||
Qy 201 AspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
|||||
Db 110706 GATCTGTCAAGAACGACGAGAAGATCTTCTGTCCCATCCCTATCCCTGCGGATACCAGG 110765
|||||
Qy 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
|||||
Db 110766 GATAATGAAGTGAATCAGCCCTCGAAGAGAGGCGGTGAAGCGGAGCAG 110825
|||||
Qy 241 AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
|||||
Db 110826 GCTGCCAGTGAAGCGGCTGGTGGAGACACTACACCGAGGTCTTCCCTTCTCTCTAC 110885
|||||
Qy 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
|||||
Db 110886 TATGAACAACCTCTGGGGCAACCTCCCGGTTCACCTCAAAAGTTACGAGCAGAGAAGTT 110945
|||||
Qy 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProProGlnVal 300
|||||
Db 110946 CCAGAGGAACTCGAGTACAGTTGGATTGTCATAGTGGTAGGAATTCACCACCCTCAAGTA 111005
|||||
Qy 301 ArgTrpTyrCys 304
|||||
Db 111006 AGTAAAAAATGT 111017
|||||

RESULT 12
AC063964 175265 bp DNA linear HTG 30-AUG-2001
LOCUS Homo sapiens chromosome 10 clone RP11-31303, WORKING DRAFT
DEFINITION AC063964
ACCESSION AC063964
VERSION AC063964.3 GI:9887616
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 175265)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175265)
```


AUTHORS
TITLE
JOURNAL
COMMENT

Smith,D.R.
Direct Submission
Submitted (22-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Aug 24, 2000 this sequence version replaced gi:8439815.

----- Genome Center

Center: Genome Therapeutics Corporation
Center code: GTC
Web site: <http://www.genomecorp.com/>
Contact: gto-seqcenter@genomecorp.com

----- Project Information

Center project name: hg198
----- Summary Statistics

Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 153700 bases at least Q40
Consensus quality: 164398 bases at least Q30
Consensus quality: 167130 bases at least Q20
Insert size: 173464; sum-of-contigs
Quality coverage: 4.1x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.

1
* 1146 1245: contig of 1145 bp in length
* 1246 1245: gap of unknown length
* 2414 1246: contig of 1169 bp in length
* 2415 2514: gap of unknown length
* 2515 3864: contig of 1350 bp in length
* 3865 3964: gap of unknown length
* 3965 5079: contig of 1114 bp in length
* 5079 5179: gap of unknown length
* 5179 6648: contig of 1469 bp in length
* 6648 7927: gap of unknown length
* 7927 8027: contig of 1179 bp in length
* 8027 9413: gap of unknown length
* 9413 10563: contig of 1386 bp in length
* 10563 10664: contig of 1051 bp in length
* 10664 11996: gap of unknown length
* 11996 12096: contig of 1332 bp in length
* 12096 13310: gap of unknown length
* 13310 13410: contig of 1214 bp in length
* 13410 15173: gap of unknown length
* 15173 15273: contig of 1763 bp in length
* 15273 17974: gap of unknown length
* 17974 21420: contig of 2701 bp in length
* 21420 21519: gap of unknown length
* 21519 24881: contig of 3346 bp in length
* 24881 24982: contig of 3362 bp in length
* 24982 31877: gap of unknown length
* 31877 49015: contig of 6895 bp in length
* 49015 49115: gap of unknown length
* 49115 66286: contig of 17039 bp in length
* 66286 66387: contig of 17171 bp in length
* 66387 86202: gap of unknown length
* 86202 86302: contig of 19815 bp in length
* 86302 132481: contig of 19815 bp in length
* 132481 132581: contig of 46179 bp in length
* 132581 175265: gap of unknown length
* 175265 42685: contig of 42685 bp in length.

FEATURES

Location/Qualifiers
1..175265
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="10"
/clone="RP11-31303"
/clone_lib="RPC1-11"
1..1145
/note="assembly_name:Contig1"
1246..2414
/note="assembly_name:Contig11"
2515..3864
/note="assembly_name:Contig12"
3965..5078
/note="assembly_name:Contig15"
5179..6647
/note="assembly_name:Contig18"
6748..7926
/note="assembly_name:Contig20"
8027..9412
/note="assembly_name:Contig23"
9513..10563
/note="assembly_name:Contig25"
10664..11995
/note="assembly_name:Contig27"
12096..13309
/note="assembly_name:Contig28"
13410..15172
/note="assembly_name:Contig29"
15273..17973
/note="assembly_name:Contig30"
18074..21419
/note="assembly_name:Contig31"
21520..24881
/note="assembly_name:Contig32"
24982..31876
/note="assembly_name:Contig33"
31977..49015
/note="assembly_name:Contig34"
49116..66286
/note="assembly_name:Contig35"
clone_end:SP6"
66387..86201
/note="assembly_name:Contig36"
86302..132480
/note="assembly_name:Contig37"
clone_end:T7"
132581..175265
/note="assembly_name:Contig38"
BASE COUNT 53132 a 34451 c 33650 g 52102 t 1930 others
ORIGIN

Alignment Scores:
Pred. No.: 7.59e-49 Length: 175265
Score: 1553.00 Matches: 302
Percent Similarity: 99.34% Conservative: 0
Best Local Similarity: 99.34% Mismatches: 2
Query Match: 22.54% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-2 (1-1320) x AC063964 (1-175265)

QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Db 162072 ATGCAGACGACAGCATAGAGCTTCTACTTCCATATCTACGCTTCTAAGAGAGAGCTAT 162131
QY 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Db 162132 TTAGCTGAACACCGACATCGGGGAAACAATGAGAGAGTGCAGCGAGCCCTCTCTCAAC 162191
QY 41 ProCysHisPheGlySerProSerGlyValaAlaGluGlyGlyGlnAspLeu 60
Db 162192 CCTTGCATTTCCGACAGTCTCTCTGGGGCCGCTCAAGAGGCGGAGGCAAGATGACCTT 162251
QY 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Db 162252 CCAGATCTTTCAGCCTTCTGAGCCAAGAGAAATTAGACGAAAGTGTCAATTTGGCAAGA 162311

QY 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaAArgLysArgLeu 100
|||||
Db 162312 CTGGCCATCAATACGACCCCTTTGGAGAGCGCATGAACATCAAGCTAGAAAACGACTT 162371

QY 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPhcCysGlnAsp 120
|||||
Db 162372 TCTCCTGATCAGATGAACACACTCACCTAAATTAAGTTTGTAGCCCTAACTTCTGCCAGGAT 162431

QY 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
|||||
Db 162432 AACCTCGAAGTCCACGACGCTCTTAAGAAAGCCCGAGGAGGCAAAAGGCCACAGTAT 162491

QY 141 CysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
|||||
Db 162492 TGTCTGAAACCCAGTCCAAAAGATATTTAAATAGGCTGCCGACTTCATTGAAGAG 162551

QY 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
|||||
Db 162552 CTATCCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACCTCGCTGCTGCAAAAAC 162611

QY 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
|||||
Db 162612 CACAAGAGTAAACTGGAATCTCAAAACAAAGTTATGCGAGGAAACACAGCTCCAGTTCTCA 162671

QY 201 AspLeuSerGluArgArgGluArgSerValProIleProIleProAlaAspThrArg 220
|||||
Db 162672 GATCTGTGAGAAACAGAGAAAGATCTTCTGTCTCCATCCCTATCCCTGCGGATACCAGG 162731

QY 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
|||||
Db 162732 GATAATGAAGTGAATCAGCCCTGGAACAGCAGCAAGCAAGAGCGCTGAAGCGGAGCAG 162791

QY 241 AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
|||||
Db 162792 GCTCCCAAGTGGGGCGGTGGTGGAGACACTACACACAGGGTCTTCCCTTCATCTCTGTAC 162851

QY 261 TyrGluGluProLeuGluGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
|||||
Db 162852 TATGAGAACCTCTGGGGCAACCTCCCGGTCTACTCAAAAGTTACGAGCAGAGAAGTT 162911

QY 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
|||||
Db 162912 CCAGAAGAACTCGATACAGTTGGATTGTCATAGTGGTAGGAATTCACCACTCAAGTA 162971

QY 301 ArgTrpTyrCys 304
|||
Db 162972 AGGTAAATAATGT 162983

RESULT 13
AC024460/c 180707 bp DNA linear HTG 28-MAR-2000
LOCUS
DEFINITION Homo sapiens clone RP11-287P14, WORKING DRAFT SEQUENCE, 22
unordered pieces.
AC024460
VERSION AC024460.2 GI:7331497
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180707)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavskiy, L.,
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M.,
Fenster, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Ollivar, T.M.,
Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,
Travers, M., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 28, 2000 this sequence version replaced gi:7108256.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7059
Center clone name: 287_P_14
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168456 bases at least Q40
Consensus quality: 174769 bases at least Q30
Consensus quality: 176914 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 178607; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1528: contig of 1528 bp in length
* 1529 1628: gap of 100 bp
* 1629 2914: contig of 1286 bp in length
* 2915 3014: gap of 100 bp
* 3015 5061: contig of 2047 bp in length
* 5062 5161: gap of 100 bp
* 5162 7974: contig of 2813 bp in length
* 7975 8074: gap of 100 bp
* 8075 11732: contig of 3658 bp in length
* 11733 11832: gap of 100 bp
* 11833 14787: contig of 2955 bp in length
* 14788 14887: gap of 100 bp
* 14888 19746: contig of 4859 bp in length
* 19747 19846: gap of 100 bp
* 19847 23870: contig of 4024 bp in length
* 23871 23970: gap of 100 bp
* 23971 28748: contig of 4778 bp in length
* 28749 28848: gap of 100 bp
* 28849 35292: contig of 6444 bp in length
* 35293 35392: gap of 100 bp
* 35393 40357: contig of 4965 bp in length
* 40358 40457: gap of 100 bp
* 40458 45936: contig of 5479 bp in length
* 45937 46036: gap of 100 bp

TITLE
JOURNAL
COMMENT

```

* 46037 52180: contig of 6144 bp in length
* 52181 52280: gap of 100 bp
* 52281 57511: contig of 5231 bp in length
* 57512 57611: gap of 100 bp
* 57612 65019: contig of 7408 bp in length
* 65020 65119: gap of 100 bp
* 65120 73479: contig of 8360 bp in length
* 73480 73579: gap of 100 bp
* 73580 81893: contig of 8314 bp in length
* 81894 81993: gap of 100 bp
* 81994 92063: contig of 10070 bp in length
* 92064 92163: gap of 100 bp
* 92164 103396: contig of 11233 bp in length
* 103397 103496: gap of 100 bp
* 103497 112935: contig of 9439 bp in length
* 112936 113035: gap of 100 bp
* 113036 126895: contig of 13860 bp in length
* 126896 126995: gap of 100 bp
* 126996 180707: contig of 53712 bp in length.
FEATURES
    source
    1..180707
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-287P14"
        /clone_lib="RPC1-II Human Male BAC"
    1..1528
        /note="assembly_fragment"
    1629..2914
        /note="assembly_fragment"
    3015..5061
        /note="assembly_fragment"
    5162..7974
        /note="assembly_fragment"
    8075..11732
        /note="assembly_fragment"
    11833..14787
        /note="assembly_fragment"
    14888..19746
        /note="assembly_fragment"
    19847..23870
        /note="assembly_fragment"
    23971..28748
        /note="assembly_fragment"
    28849..35292
        /note="assembly_fragment"
    35393..40357
        /note="assembly_fragment"
    40458..45936
        /note="assembly_fragment"
    46037..52180
        /note="assembly_fragment"
    52281..57511
        /note="assembly_fragment"
    57612..65019
        /note="assembly_fragment"
    65120..73479
        /note="assembly_fragment"
    73580..81893
        /note="assembly_fragment"
    81994..92063
        /note="assembly_fragment"
    92164..103396
        /note="assembly_fragment"
    103497..112935
        /note="assembly_fragment"
    clone_end:77
    vector_side:left"
    113036..126895
        /note="assembly_fragment"
    126996..180707
        /note="assembly_fragment"
    clone_end:SP6
    vector_side:left"

```

```

BASE COUNT  55951 a 33545 c 34680 g 54429 t 2102 others
ORIGIN
Alignment Scores:
Pred. No.:      7.85e-49      Length:      180707
Score:          1553.00      Matches:    302
Percent Similarity: 99.34%    Conservatives: 0
Best Local Similarity: 99.34%  Mismatches:  2
Query Match:    22.54%      Indels:     0
DB:             2           Gaps:       0

```

US-09-818-990B-2 (1-1320) x AC024460 (1-180707)

```

QY 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
|||||
Db 154380 ATCAAGACGACAGCATAGAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 154321
QY 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgLaGluProSerSerAsn 40
|||||
Db 154320 TTAGCTGAACACGACATCGGGAAACAATGAGAGGAGTCGAGCGGAGCCCTCTCCAAC 154261
QY 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeu 60
|||||
Db 154260 CCTTGGCCATTTTCGGCAGTCCTTCTGGGGCCGCTGAAGGAGCGGAGCCCAAGATGACCTT 154201
QY 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
|||||
Db 154200 CCAGATCTTTCAGCCCTTCTGAGCCCAAGAAAGATTAGACCAAAAGTGCAATTTGGCAAGA 154141
QY 81 LeuAlaIleAsnTyrAspProLeuGluLeuLysAlaAspGluThrGlnAlaArgLysArgLeu 100
|||||
Db 154140 CTGGCCATCAATTACGACCCCTTTGGAGAGGCAGATGAAACTCAAGCTAGAAACGACTT 154081
QY 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
|||||
Db 154080 TCTCTGATCAGATGAACACTCACCTTAATTTAAGTTTTCAGCTTCTTCTGCCAGAT 154021
QY 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
|||||
Db 154020 AACCTCGAAGTCCACACCTCTTAAGAAAGCCCGGAGGCGGCAAGGCGCACAGTAT 153961
QY 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
|||||
Db 153960 TGTTCCTGAACCCAGTCCAAAAAAGTATTTTAAATAAGGCTGCGGACTTCATTGAAGAG 153901
QY 161 LeuSerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsn 180
|||||
Db 153900 CTATCTCTCCCTTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCTGCTGCTGCAAAAAC 153841
QY 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerSerPheSer 200
|||||
Db 153840 CACAAGAGTAAACTGGAACTCAAAACAAGTTATGCGAGGAAACACAGCTCCAGTTTCTCA 153781
QY 201 AspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
|||||
Db 153780 GATCTGTCAAGAAAGACGAGAAAGATCTTCTGTTCCTATCCCTATCCCTGCGGATACCAGG 153721
QY 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGlnGlnLysArgArgGluAlaGluGln 240
|||||
Db 153720 GATAATGAAGTGAATCACGCCCTTGGACAGCAGGAGGCGGCTGAAGCGGAGCAG 153661
QY 241 AlaAlaSerGluAlaAlaGlyLysValThrProGlySerSerProSerSerLeuTyr 260
|||||
Db 153660 GCTGCCAGTGGAGGCGGCTGGTGAGACACACACACAGGGTCTTCCCTCTCATCTCTGTAC 153601
QY 261 TyrGluGluProLeuGlyGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
|||||
Db 153600 TATGAAGAACTCTGGGGCAACCTCCCGGTTTCACTCAAAAGTTTACGAGCAGAGAGATT 153541
QY 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
|||||
Db 153540 CCAGAAGGAACTCGAGTACAGTTGGATTGATGTAGTAATTCACCACTCAAGTA 153481

```

Qy 301 ArgTrpTyrCys 304

Db 153480 AGCTAAAAATGCT 153469

RESULT 14

BC013867

LOCUS BC013867 Homo sapiens, similar to palladin, clone MGC:10248 IMAGE:3845400, mRNA, complete cds.

DEFINITION

ACCESSION BC013867

VERSION BC013867.1

KEYWORDS MGC.

SOURCE Homo sapiens.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (07-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-x@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>contact: amadan@systemsbiology.org

Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia

Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 14 Row: m Column: 24

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 4689129.

FEATURES

source

```

1..1664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:10248 IMAGE:3845400"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
387..1544
/codon_start=1
/product="Similar to palladin"
/protein_id="AAH13867.1"
/db_xref="GI:15530191"
/translation="MAPFFEMKLHKYIFEGMPVFTFCRVAGNPKPKIYWFKDKQIS
PKSDHYTIQRLDGTCSLHTTASTLDDGNVTIMANPOGRISCTGRMLVMQAVNQRGR
SPRSPGHVPRPRSRSDGNEPIQERFRPHFLOAPGDLTVQEGKLCRMDCKV
SGLTPDLSQLDGKVPKPSAHKMLVRENGVHSLIEPVTSRDAGLYTCIATNRAGQ
NSFSLVYAKAKAHAPPPVIEKLTQNTGVADGVPVLECRVLGVPPPPQIFWKXENESL
THSTDVSMQDNHGYICLLIQGATKEDAGWTVTSKNEAGIVSCVTRALDVTYQWQQ
SQSTPKPKVPSASRYAALSDQGLDKAAAFQPEANPSPHSLTFLNLTALVESEDL"
BASE COUNT 514 a 397 c 406 g 347 t
ORIGIN

```

Alignment Scores:

Pred. No.:	3,76e-48	Length:	1664
Score:	1476.00	Matches:	278
Percent Similarity:	70.51%	Conservative:	83
Best Local Similarity:	54.30%	Mismatches:	107
Query Match:	21.42%	Indels:	44
DB:	9	Gaps:	4

US-09-818-990B-2 (1-1320) x BC013867 (1-1664)

Qy	852	AlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSerProGlnProVal	871
Db	9	GCCAAAGGTTGTCACCCCGCAGGATTTCCAAAGAAAGCCAGTAGACTGTAGTAGACC	68
Qy	872	AsnAspAsnIleAArgGluThrLysAsnAlaValIleAArgPLeuGlyLysLysIle	891
Db	69	TCCGATGAGGAAATTCAGGCACAAGGATGCTGTTATTCAAGACCTTGAACGAAACTT	128
Qy	892	ThrPheSerAspValAArgProAsn	899
Db	129	CGCTTCAAGGAGGACCTCCTGGAACAATGGCCAGCGAGTTAACATACGAAGAAGAAATG	188
Qy	899		899
Db	189	GCTCGTCGACTGCTAGGTGCTGACAGTCGCAACTGCTTTTAATATTTCAGGAGCCAGAGAG	248
Qy	900	-----GlnGlnGluTyrLysIleSerSerPheGluGlnAArgLeuMetAsnGluIle	916
Db	249	GAACACAGCTAATCAGGAATACAAAGTCTCCAGCTGTGAACAGAGACTCATCAGTGAATA	308
Qy	917	GluPheAArgLeuGluAArgThrProValAspGluSerAspAspGluIleGlnHisAspGlu	936
Db	309	GAGTACAGGCTAGAAAGTCTCCTGTGGATGAATCAGTGTGATGAAGTTCAGTATGGAGAT	368
Qy	937	IleProThrGlyLysCysIleAlaProIlePheAspLysAArgLeuLysHisPheAArgVal	956
Db	369	GTGCCTGTGNAATGGAATGGACCACTTTTCAGATGAAGCTGAACATCATTACAAGATC	428
Qy	957	ThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysVal	976
Db	429	TTTCAGGGAATGTCAGTAACTTTTACATGTAGAGTGTGGAAATCCAAAGCCCAAGATC	488
Qy	977	TyrTrpPheLysAspGlyLysGlnIleSerLysAArgAsnGluHisCysLysMetAArgArg	996
Db	489	TATGGTTTAAAGATGGGAAGCAGATCTCTCCAAGAGTGTATCACTACACATTTCAAAGA	548
Qy	997	GluGlyAspGlyThrCysSerLeuHisIleGluSerThrSerAspAspGlyAsn	1016
Db	549	GATCTCATGGGACCTGCTCCCTCCATACCACAGCCCTACCTAGATGATGATGGGAAT	608
Qy	1017	TyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetVal	1036
Db	609	TATACAAATATGCTGCAAAACCTCAGGGCGCATCATGTTGTACTGGACGGCTAATGTA	668
Qy	1037	GlnSerLeuProIleAArgSerAArgLeuThrSerAlaGlyGlnSerHis	1052
Db	669	CAGGCTGTCAACCAAGAGGTGGAAGTCCCGGTCTCCCTCAGGCCCATCTCTATGTCAGA	728
Qy	1053	-----ArgGlyAArgSerAArgValGlnGluAArgAspLysGluProLeuGlnGluAArgPhe	1070
Db	729	AGGCCTCGTTCTAGATCAAGGGACAGTGGAGACCAAAATGAACCAATTCAGGACGGATTC	788
Qy	1071	PheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyAArgLeuCys	1090
Db	789	TTTCAGACCTCACTTCTTCAGGCTCCTGGAGATCTGCTGTTCAGGAAGGAAAACTCTGC	848
Qy	1091	ArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGly	1110
Db	849	AGAATGGACTGCAAGTCAGTGGGTTCACCAACCCGATCTTAAGCTGGCAACTAGATGGA	908
Qy	1111	GlnProValLeuProAspAlaSerHisLysMetLeuValAArgGluThrGlyValHisSer	1130
Db	909	AAGCCCGTACCCCTGACAGTGTCTACAGATGCTGGTGGTGAACGGGGTCCACTCT	968
Qy	1131	LeuLeuIleAspProLeuThrGlnAArgAspAlaGlyThrTyrLysCysIleAlaThrAsn	1150
Db	969	CTGATCATAGGCGAGTCACGTCACGTGATGCCGCGCATCTACACATGTATAGCTACCAAC	1028
Qy	1151	LysThrGlyGlnAsnSerPheSerLeuGluSerValValAlaLysGluValLysLys	1170
Db	1029	CGAGCAGGACAGAACTCATTCAGCCTGAGCTTGTGGTTCGCTTAAGAAAGACACAAA	1088

```
QY 1171 AlaProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArg 1190
Db 1089 CCCCTGTGTATTATGAGAGCTCCAAAACACAGAGGTTCTCTGATGGGTACCCAGTGGCG 1148
QY 1191 LeuGluCysArgValIleGlyMetProProValPheTyrTrpLysLysAspAsnGlu 1210
Db 1149 CTGGAATGTCGTATTGGGAGTGCACCAACACCTCAGATATTTTGAAGAAGAAATGAA 1208
QY 1211 ThrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCys 1230
Db 1209 TCACCTCACTACAGCACTGACCGAGTGAGCATGCACAGGACAAACACGCTACATCTGC 1268
QY 1231 LeuLeuIleGlnProLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsn 1250
Db 1269 CTGCTCATTCAGGAGGCCACAAAAGAAGATGCTGGGTATATCTGTGTGAGCAAGAAT 1328
QY 1251 GluAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHis--- 1269
Db 1329 GAAGCAGGGATTGTGCTGTACTGCCAGCTGGACGCTTTACACCCAGTGGCATCAGCAG 1388
QY 1270 ---GlnIleProProMetSerValArgProSerGlySerArgTyrGlySerLeuThr 1288
Db 1389 TCACAGAGCACCAAGCAAAAAGTACGGCCCTCAGCCAGTGCCTATGCAGCACATTTCG 1448
QY 1289 SerLysGlyLeuAspIlePheSerAlaPheSerMetGluSerThrMetValTyrSer 1308
Db 1449 GACCAGGGACTAGACATCAAGCAGCGTTCACACCT---GAGGCCAACCCATCTCACCTG 1505
QY 1309 CysSerSerArgSerValValGluSerAspGluLeu 1320
Db 1506 ACACCTGAATACTGCTTGGTGAAGTGGAGGACCTG 1541

RESULT 15
AF205078
LOCUS
DEFINITION
Mus musculus strain Swiss Webster/NIH actin-associated protein
palladin mRNA, partial cds.
ACCESSION
AF205078
VERSION
AF205078.1 GI:9828172
KEYWORDS
Mus musculus.
SOURCE
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1580)
Parast,M.M. and Otey,C.A.
Characterization of palladin, a novel protein localized to stress
fibers and cell adhesions
J. Cell Biol. 150 (3), 643-656 (2000)
20391984
PUBMED
10931874
REFERENCE
2 (bases 1 to 1580)
Parast,M.M. and Otey,C.
Direct Submission
Submitted (16-NOV-1999) Department of Cell and Molecular
Physiology, University of North Carolina-Chapel Hill, 52 MSRB Bldg,
Campus Box 7545, Chapel Hill, NC 27599, USA
JOURNAL
Location/Qualifiers
FEATURES
1..1580
/organism="Mus musculus"
/strain="Swiss Webster/NIH"
/db_xref="taxon:10090"
/dev_stage="11.5 day embryo"
<1..1479
/note="Involved in cytoskeletal assembly"
/codon_start=1
/product="actin-associated protein palladin"
/protein_id="AAG00078.1"
/db_xref="GI:9828173"
/translation="ASDEIQTGKDAVDQLERLRFKEDLLNNGSPRLTYBERMARR
LIGASANVFNOEPEETAAAGYKVSCEQLRISEIYRLERSPVDESGDVQPDV
PVENATAPFEEKLKHKIFEGMPVFTFCRVAGNPKPIYFKDKGKQISPKSDHYTIQ
RDLGTCSLHTTASTLDDNGNTIMAAANPQGRVSVCTGRLMVQAVNQRGRSPSPSGHP
source
CDS
```

```
3'UTR
BASE COUNT 415 a 437 c 427 g 301 t
ORIGIN
Alignment Scores:
Pred. No.: 1,44e-47 Length: 1580
Score: 1460.00 Matches: 274
Percent Similarity: 71.54% Conservative: 78
Best Local Similarity: 55.69% Mismatches: 96
Query Match: 21.19% Indels: 44
Db: 10 Gaps: 5
US-09-818-990b-2 (1-1320) x AF205078 (1-1580)
QY 872 AsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeuGlyLysIle 891
Db 4 TCTGATGAGGAGATTCAAGCCACAAAGGATGCTGTATCCAAAGACCTGGCAACGAAGCTT 63
QY 892 ThrPheSerSerValArgProAsn----- 899
Db 64 CGCTTCAAGGAGGACCTTCTGAACAATGGCCAACCGAGGCTAACTATGAGAAAGAATG 123
QY 899 ----- 899
Db 124 GCTCGCCGCTGCTTGGAGCCGACAGCGCAACAGCTCTTCAACATCCAGAGCCAGAGAA 183
QY 900 -----GlnGlnGluTyrLysIleSerSerPheGlnGlnArgLeuMetAsnGluIle 916
Db 184 ACGGCACCCCAATCAGGAGTACAAGTCTCTAGCTGCGAGCAGAGCTGATTAGCGAGATT 243
QY 917 GluPheArgLeuGluArgThrProValAspGluSerAspGluIleGlnHisAspGlu 936
Db 244 GACTACCGGCTGGAGCGCTCCCTGTGGATGAGTGGGAGACGAGGTGAGGATCCAGAT 303
QY 937 IleProThrGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgVal 956
Db 304 GTCCTGTGGAGAACCAACAGCTCCCTTTGAGATGAAGTGAACACTACAAGAT 363
QY 957 ThrGluGlySerProValThrPheThrCysLysIleValGlyIleProValProLysVal 976
Db 364 TTTGAGGGGATGCGGTGACTTTCACGTGCGAGTGGCTGGGAATCCAAAGCCAAAGATC 423
QY 977 TyrTrpPheLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArg 996
Db 424 TATTGGTTTAAAGATGGGAAGCAGATTCTCCGAAGAGCGCATCACTACACCATTCAGAGA 483
QY 997 GluGlyAspGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspAspGlyAsn 1016
Db 484 GACCTTGTGGACCTGCTCTCTCCACACACCGCCCTCTACCCCTAGACGATGGGAAC 543
QY 1017 TyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetVal 1036
Db 544 TACACCATCATGGCTGCCAACCCCTCAGGGTGGCTGAGTCACTGTACAGGAAGGCTAATGGTA 603
QY 1037 GlnSerLeuProIleArgSerArg-----LeuThrSerAlaGlyGlnSerHisArgGly 1054
Db 604 CAGGCTGTCAACCAAGAGGCGCGAGTGGAGATGAAACGAGGCCATCTCATGCCAGA 663
QY 1055 ArgSerArgValGlnGluArgAsp-----LysGluProLeuGlnGluArgPhe 1070
Db 664 AGGCTCGCTCTCGATCACGGAGCAGTGGAGATGAAACGAGGCCATCTCAGGAGCATTC 723
QY 1071 PheArgProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCys 1090
Db 724 TTCAGACCTCACCTCCCTGAGGCTCTCGAGACCTGACCGTTTCAGGAAGGCAAGCTCTCG 783
QY 1091 ArgLeuAspCysLysValSerGlyLeuProProGluLeuThrTrpLeuLeuAsnGly 1110
```


GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: November 30, 2002, 17:03:50 ; Search time 73 Seconds
(without alignments)
14467.732 Million cell updates/sec
Title: US-09-818-990B-1
Perfect score: 1316
Sequence: 1 atgcaagacagacataga.....tggagagtgtgaactttaa 3963

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters: -DEV=xlh
-MODEL=frame_n2p.model
-Q=/cgn2_1/USPTO.spool/US09818990/runat_26112002_093221_22022/app_query.fasta_1.4103
-DB=A_Geneseq_101002 -QPMF=fastan -SUFFIX=oligo.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09818990_@CEN_1_1_91-runat_26112002_093221_22022 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	494	37.5	507	22	AAB92617	Human protein sequ
2	261	19.8	264	22	ABB10330	Human cDNA SEQ ID
3	225	17.1	408	22	ABG10611	Novel human diagno
4	179	13.6	261	22	AAU17989	Human immunoglobul
5	113	8.6	159	22	ABBI0509	Human cDNA SEQ ID
6	113	8.6	159	22	ABU18053	Human immunoglobul
7	81	6.2	130	21	ABF09582	Human ORFX ORF639
8	33	2.5	50	23	ABP09582	Human ORFX protein
9	19	1.4	433	21	ABF07380	Human COLX ORF2526
10	19	1.4	544	21	AA53464	Human colon cancer
11	16	1.2	43	22	AAU01767	Human secreted pro
12	16	1.2	44	22	AAU01765	Human secreted pro
13	16	1.2	49	22	AAU01763	Human secreted pro
14	14	1.1	104	22	AAU06795	Human foetal prote
15	14	1.1	162	23	ABP07380	Human ORFX protein
16	11	0.8	776	22	ABG05279	Novel human diagno
17	11	0.8	776	23	ABH95590	Human angiogenesis
18	11	0.8	776	23	ABH4984	Human PRO4527 prot
19	11	0.8	777	22	ABG20257	Novel human diagno
20	10	0.8	881	22	ABG05280	Novel human diagno
21	10	0.8	881	22	ABG20258	Novel human diagno
22	10	0.8	1892	21	AAG50497	Arabidopsis thalia
23	10	0.8	1940	21	AAG50496	Arabidopsis thalia
24	10	0.8	1944	21	AAG50495	Arabidopsis thalia
c 25	9	0.7	223	23	ABG60118	Human DTHP polype
c 26	9	0.7	3502	22	ABH58382	Drosophila melanog
27	8	0.6	14	21	AAV58492	HHV8 ORF 65-derive
28	8	0.6	17	22	ABH42312	Peptide #9818 enco
29	8	0.6	17	22	AAM63198	Human brain expres
30	8	0.6	17	22	AAM76012	Human bone marrow
31	8	0.6	17	22	AAW36120	Peptide #10157 enc
32	8	0.6	31	21	AAB38238	Human secreted pro
c 33	8	0.6	31	22	AAB20046	muO-conopeptide G2
34	8	0.6	39	22	AAU01766	Human secreted pro
c 35	8	0.6	46	22	ABH38283	Peptide #5789 enco
c 36	8	0.6	46	22	ABH23464	Protein #5463 enco
c 37	8	0.6	46	22	AAW58906	Human brain expres
c 38	8	0.6	46	22	AAW71428	Human bone marrow
c 39	8	0.6	46	22	AAW19077	Peptide #5511 enco
c 40	8	0.6	46	22	AAW31716	Peptide #5753 enco
41	8	0.6	46	22	AAU01764	Human secreted pro
c 42	8	0.6	46	23	ABG41235	Human peptide enco
c 43	8	0.6	48	22	ABH70089	Drosophila melanog
44	8	0.6	51	22	ABH39204	Peptide #6710 enco
45	8	0.6	51	22	ABH24059	Protein #6058 enco
46	8	0.6	51	22	AAW59864	Human brain expres
47	8	0.6	51	22	AAW72453	Human bone marrow
48	8	0.6	51	22	AAW19588	Peptide #6022 enco
49	8	0.6	51	22	AAW32702	Peptide #6739 enco
50	8	0.6	51	23	ABG42273	Human peptide enco
51	8	0.6	52	22	AAU44417	Propionibacterium
52	8	0.6	58	22	AAU01761	Human secreted pro
53	8	0.6	63	22	AAW74567	Human colon cancer
54	8	0.6	66	22	AAU48208	Propionibacterium
55	8	0.6	74	22	ABH17295	Human nervous syst
c 56	8	0.6	77	22	AAU53862	Propionibacterium
c 57	8	0.6	80	22	AAW20056	muO-conopeptide G2
58	8	0.6	85	22	AAU57363	Propionibacterium
59	8	0.6	91	23	ABP04844	Human ORFX protein
c 60	8	0.6	93	20	AAU12884	Human 5' EST secre
c 61	8	0.6	93	21	AAV65059	Human 5' EST relat
62	8	0.6	98	21	AAV55519	Arabidopsis thalia
63	8	0.6	108	20	AAV19794	B. burgdorferi ant
64	8	0.6	111	22	AAU49960	Propionibacterium
c 65	8	0.6	114	22	ABH11259	Human kinase homol
66	8	0.6	119	22	AAU40112	Propionibacterium
c 67	8	0.6	119	22	AAU62831	Propionibacterium
68	8	0.6	129	20	AAV19793	B. burgdorferi ant

69	8	0.6	133	22	AA093627	Human polypeptide,	142	8	0.6	362	23	AB092431	Herbicidally activ
c 70	8	0.6	134	23	ABP07102	Human ORFX protein	c 143	8	0.6	368	21	AB026843	Sakuranetin synth
c 71	8	0.6	141	20	AA036045	Extended human sec	144	8	0.6	381	21	AA039847	Arabidopsis thalia
72	8	0.6	142	21	AA093408	Amino acid sequenc	c 145	8	0.6	389	21	AA030620	Murine beta-1,3-ga
73	8	0.6	146	23	ABP42439	Human ovarian anti	c 146	8	0.6	392	12	AA014662	El-beta subunit of
c 74	8	0.6	149	22	AA060529	Propionibacterium	c 147	8	0.6	397	21	AA030619	Human beta-1,3-gal
75	8	0.6	158	21	AA055518	Arabidopsis thalia	c 148	8	0.6	437	22	AA082933	S. epidermidis ope
76	8	0.6	165	21	AA060643	Arabidopsis thalia	149	8	0.6	464	22	AB065760	Drosophila melanog
77	8	0.6	173	21	AA055517	Arabidopsis thalia	150	8	0.6	501	19	AA082571	Human Bp1 DNA. H
78	8	0.6	173	22	AA038789	Novel subtilisin h	151	8	0.6	519	23	AA093044	Arabidopsis transc
79	8	0.6	173	22	AA038805	Novel subtilisin h	c 152	8	0.6	595	22	AA036514	Pseudomonas aerugi
80	8	0.6	173	22	AA038859	Novel subtilisin h	c 153	8	0.6	603	22	AB064542	Drosophila melanog
81	8	0.6	174	22	AB014566	Novel human diagno	c 154	8	0.6	605	23	ABP39580	Staphylococcus epi
82	8	0.6	176	21	AA093412	Amino acid sequenc	155	8	0.6	646	21	AA057307	P. aeruginosa matu
c 83	8	0.6	178	22	AA012073	Dendritic cell (DC	156	8	0.6	670	21	AA057314	P. aeruginosa cera
c 84	8	0.6	179	22	AB021767	Novel human diagno	157	8	0.6	745	23	AB057319	Mouse ischaemic co
c 85	8	0.6	192	19	AA074885	Human secreted pro	158	8	0.6	746	20	AA028882	Human Ext-1 protei
c 86	8	0.6	192	21	AA091681	Human secreted pro	159	8	0.6	748	23	AB028874	Herbicidally activ
c 87	8	0.6	198	23	AB048986	Listeria monocytog	160	8	0.6	761	22	AB058572	Drosophila melanog
c 88	8	0.6	211	21	AA023517	Arabidopsis thalia	161	8	0.6	783	22	AB040903	Novel human diagno
89	8	0.6	217	21	AA060643	Arabidopsis thalia	162	8	0.6	792	22	AB064798	Drosophila melanog
90	8	0.6	218	21	AA024454	Human secreted pro	c 163	8	0.6	793	22	AB065720	A. gossypii AG011
91	8	0.6	219	21	AA023516	Arabidopsis thalia	164	8	0.6	848	21	AA070955	Arabidopsis SERK 1
c 92	8	0.6	221	22	AA045702	Human 7TM clone HD	c 165	8	0.6	996	22	AA090093	C glutamicum prote
c 93	8	0.6	225	21	AA094449	Human inflammation	166	8	0.6	1051	12	AA014117	Hamster GAP b3 pro
c 94	8	0.6	225	22	AA065272	Human cell surface	167	8	0.6	1250	21	AA091279	Group B Streptococ
95	8	0.6	229	21	AA060643	Arabidopsis thalia	168	8	0.6	1252	23	ABP26711	Streptococcus poly
96	8	0.6	229	21	AA039849	Arabidopsis thalia	169	8	0.6	1252	23	ABP29749	Streptococcus poly
c 97	8	0.6	239	22	AA074584	Human HAERBs isof	c 170	8	0.6	1285	21	AA043359	Human ORFX ORF3123
c 98	8	0.6	240	21	AA026845	Sakuranetin synth	171	8	0.6	1337	16	AA085203	huBP-1. Homo sep
c 99	8	0.6	242	21	AA073495	Human secreted pro	172	8	0.6	1510	20	AA010131	Glutamine-Oxogluta
c 100	8	0.6	248	20	AA015225	Human receptor pro	173	8	0.6	1510	22	AA089953	C glutamicum prote
c 101	8	0.6	248	20	AA036046	Extended human sec	174	8	0.6	1510	22	AA079669	Corynebacterium gl
c 102	8	0.6	248	21	AA091531	Human secreted pro	175	8	0.6	1527	20	AA043543	A human MPR-relate
c 103	8	0.6	248	22	AA012072	Dendritic cell (DC	176	8	0.6	1527	23	AA091309	Human protein NOV1
c 104	8	0.6	248	22	AA070469	Human HAERBs-iso	177	8	0.6	1528	18	AA033363	Human multidrug re
c 105	8	0.6	248	23	AB090341	Human polypeptide	178	8	0.6	1953	23	AA084351	Protein MYLK diffe
c 106	8	0.6	249	22	AA074583	Human HAERBs isof	c 179	8	0.6	2023	22	AB063487	Drosophila melanog
c 107	8	0.6	250	20	AA048505	Human breast tumou	c 180	8	0.6	2053	22	AA003501	Human protein kina
c 108	8	0.6	257	21	AA058419	Lung cancer associ	c 181	8	0.6	2053	23	AA024079	Human MPR protein
c 109	8	0.6	259	22	AA030673	Gopher apple oleoy	c 182	8	0.6	2054	23	AA024150	Human kinase (PKIN
110	8	0.6	271	21	AA014830	Bacillus lentus I-	c 183	8	0.6	2547	19	AA081511	Drosophila fat fac
c 111	8	0.6	273	21	AA091680	Human secreted pro	c 184	8	0.6	2559	23	AB057181	Mouse ischaemic co
c 112	8	0.6	273	22	AB011989	Human secreted pro	c 185	8	0.6	2587	22	ABG20102	Novel human diagno
c 113	8	0.6	273	22	AA025809	Human protein sequ	c 186	8	0.6	2644	22	ABG20103	Novel human diagno
c 114	8	0.6	273	22	AA075619	Human colon cancer	c 187	8	0.6	2778	22	AB058683	Drosophila melanog
c 115	8	0.6	279	22	AA093492	Human polypeptide,	c 188	8	0.6	3594	23	AA020147	Mouse C3b/C4b comp
c 116	8	0.6	295	23	AB048857	Listeria monocytog	189	8	0.5	9	17	AA000714	Human Stat 5 fragm
c 117	8	0.6	298	22	AA041987	Human polypeptide	190	8	0.5	10	20	AA045955	Immunogenic peptid
c 118	8	0.6	299	22	ABG09593	Novel human diagno	c 191	7	0.5	10	22	AA071919	Random 10-mer pept
c 119	8	0.6	305	23	AA016509	Rice caffeic acid	192	7	0.5	10	23	ABP30935	HLA-A2 10-mer bind
120	8	0.6	309	21	AA023515	Arabidopsis thalia	c 193	7	0.5	11	21	AA009394	Hepatitis GB virus
c 121	8	0.6	335	22	ABG06645	Novel human diagno	c 194	7	0.5	14	22	AA068142	Peptide derived fr
122	8	0.6	339	21	AA054254	Human pancreatic c	c 195	7	0.5	14	22	AA068143	Peptide derived fr
123	8	0.6	352	18	AA014795	Mangosteen Class I	c 196	7	0.5	14	22	AA068174	Peptide derived fr
124	8	0.6	352	18	AA009298	Mangosteen Class I	c 197	7	0.5	14	22	AA068175	Peptide derived fr
125	8	0.6	352	20	AA087471	Mangosteen acyl AC	c 198	7	0.5	14	22	AA068176	Peptide derived fr
126	8	0.6	352	20	AA087472	Mangosteen acyl AC	c 199	7	0.5	14	22	AA068177	Peptide derived fr
127	8	0.6	352	20	AA098150	Mangosteen acyl AC	200	7	0.5	15	18	AA038981	Peptide resembling
128	8	0.6	352	20	AA087469	Mangosteen acyl AC	201	7	0.5	15	22	AA068653	Human cytomagalovi
129	8	0.6	352	20	AA087466	Mangosteen C18:1 a	202	7	0.5	15	22	AA068654	Human cytomagalovi
130	8	0.6	352	21	AA082090	Mangosteen FataA-ty	203	7	0.5	15	22	AA068655	Human cytomagalovi
131	8	0.6	352	23	AA096742	Mangosteen FataA-ty	204	7	0.5	18	22	AA089036	HIV gp120 protein
132	8	0.6	356	22	AA025563	Human G Protein-Co	205	7	0.5	18	22	AA089037	HIV gp120 protein
133	8	0.6	362	13	AA029172	Brassica thioester	206	7	0.5	18	22	AA089038	G-protein coupled
134	8	0.6	362	15	AA044948	Brassica campestri	207	7	0.5	20	20	AA039286	Human MDR1-p glyco
135	8	0.6	362	16	AA074150	Brassica acyl thio	c 208	7	0.5	21	18	AA034050	Fusin receptor ant
136	8	0.6	362	19	AA044333	Brassica campestri	209	7	0.5	21	18	AA034004	ATP-binding cassel
137	8	0.6	362	20	AA028638	Brassica napus acy	c 210	7	0.5	21	22	AA040308	ATP-binding cassel
138	8	0.6	362	20	AA087467	Brassica rapa C18:	211	7	0.5	21	23	ABP30910	OE8 antibody epit
139	8	0.6	362	21	AA039848	Arabidopsis thalia	212	7	0.5	21	23	AA089534	Insulin/insulin-li
140	8	0.6	362	21	AA033760	An acyl-ACP thioes	213	7	0.5	23	20	AA039241	G-protein coupled
141	8	0.6	362	21	AA053761	An acyl-ACP thioes	214	7	0.5	24	22	AA060446	Human brain expres

c 215	7	0.5	24	22	AAW73087	Human bone marrow	c 288	7	0.5	52	22	ABB29542	Peptide #2193 enco
c 216	7	0.5	24	22	AAW33306	Peptide #7343 enco	c 289	7	0.5	52	22	ABB34727	Peptide #2233 enco
c 217	7	0.5	24	22	ABG42933	Human peptide enco	c 290	7	0.5	52	22	ABB20139	Protein #2138 enco
c 218	7	0.5	25	22	ABG40984	Peptide #8490 enco	c 291	7	0.5	52	22	AAW81029	Human haematologic
c 219	7	0.5	25	22	AAW61844	Human brain expres	c 292	7	0.5	52	22	AAW81262	Human haematologic
c 220	7	0.5	25	22	AAW34759	Peptide #8796 enco	c 293	7	0.5	52	22	AAW55524	Human brain expres
c 221	7	0.5	26	10	AAW91262	Inhibin 18 kD chai	c 294	7	0.5	52	22	AAW67908	Human bone marrow
c 222	7	0.5	26	23	AAU93966	Bovine inhibin alp	c 295	7	0.5	52	22	AAW15725	Peptide #2159 enco
c 223	7	0.5	28	22	AAW51686	PAM related peptid	c 296	7	0.5	52	22	AAW28230	Peptide #2267 enco
c 224	7	0.5	29	9	AAW81907	N-terminal of inhi	c 297	7	0.5	52	22	AAW03459	Peptide #2141 enco
c 225	7	0.5	30	22	ABW37575	Peptide #5081 enco	c 298	7	0.5	52	23	ABG37444	Human peptide enco
c 226	7	0.5	30	22	AAW89359	Human immune/haema	c 299	7	0.5	52	23	ABP10517	Human ORFX protein
c 227	7	0.5	32	22	AAW92074	Inhibin peptide SE	c 300	7	0.5	53	22	ABG22991	Novel human diagno
c 228	7	0.5	32	22	AAW92078	Inhibin peptide SE	c 301	7	0.5	53	22	ABW29253	Peptide #1904 enco
c 229	7	0.5	33	12	AAW12087	N-terminal of 18 k	c 302	7	0.5	53	22	ABW34418	Peptide #1924 enco
c 230	7	0.5	33	22	AAW92075	Inhibin peptide SE	c 303	7	0.5	53	22	ABW19830	Protein #1829 enco
c 231	7	0.5	33	22	AAW92079	Inhibin peptide SE	c 304	7	0.5	53	22	AAW55208	Human brain expres
c 232	7	0.5	34	15	AAW66852	Rice mitochondrial	c 305	7	0.5	53	22	AAW67603	Human bone marrow
c 233	7	0.5	34	15	AAW74754	Rice mitochondrial	c 306	7	0.5	53	22	AAW15410	Peptide #1844 enco
c 234	7	0.5	34	21	AAW28738	Human secreted pro	c 307	7	0.5	53	22	AAW27899	Peptide #1936 enco
c 235	7	0.5	34	23	ABG66875	Human prostate spe	c 308	7	0.5	53	22	AAW03172	Peptide #1854 enco
c 236	7	0.5	35	21	AAW45245	Human secreted pro	c 309	7	0.5	53	22	ABG37193	Human peptide enco
c 237	7	0.5	35	21	AAW56759	Arabidopsis thalia	c 310	7	0.5	54	22	AAU59008	Propionibacterium
c 238	7	0.5	36	22	ABG14004	Novel human diagno	c 311	7	0.5	54	22	AAU67102	Propionibacterium
c 239	7	0.5	37	21	AAW13519	Human glycine tran	c 312	7	0.5	54	22	AAU14366	Human novel protei
c 240	7	0.5	37	21	AAW56758	Arabidopsis thalia	c 313	7	0.5	55	20	AAW19563	Amino acid sequenc
c 241	7	0.5	38	22	AAW89067	Human immune/haema	c 314	7	0.5	55	21	AAW35586	Arabidopsis thalia
c 242	7	0.5	39	20	AAW74414	HPM091 protein se	c 315	7	0.5	55	22	ABG16736	Novel human diagno
c 243	7	0.5	39	22	ABW44117	Peptide #11623 enc	c 316	7	0.5	56	22	AAU41382	Propionibacterium
c 244	7	0.5	39	22	ABW27004	Protein #9003 enco	c 317	7	0.5	56	23	ABP32402	Human ORF1375 prot
c 245	7	0.5	39	22	AAW65137	Human brain expres	c 318	7	0.5	57	21	AAW57357	Arabidopsis thalia
c 246	7	0.5	39	22	AAW77844	Human bone marrow	c 319	7	0.5	57	21	AAW60512	Arabidopsis thalia
c 247	7	0.5	39	22	AAW21748	Peptide #8182 enco	c 320	7	0.5	57	23	ABP01126	Human ORFX protein
c 248	7	0.5	39	22	AAW38064	Peptide #12101 enc	c 321	7	0.5	58	22	AAU54686	Propionibacterium
c 249	7	0.5	39	23	ABG46872	Human peptide enco	c 322	7	0.5	58	22	AAU55851	Propionibacterium
c 250	7	0.5	40	23	ABP29090	Streptococcus poly	c 323	7	0.5	58	22	ABG28606	Novel human diagno
c 251	7	0.5	41	13	AAW25795	[GLu2,13,22, Leu12	c 324	7	0.5	58	22	AAW75464	Human colon cancer
c 252	7	0.5	42	20	AAW36254	Human secreted pro	c 325	7	0.5	58	22	AAU00689	Thymosin-beta-10-1
c 253	7	0.5	42	21	AAW56198	Human secreted pro	c 326	7	0.5	58	23	AAW23797	Thymosin-beta-10-1
c 254	7	0.5	42	22	ABW41833	Peptide #9339 enco	c 327	7	0.5	59	22	AAU45252	Propionibacterium
c 255	7	0.5	42	22	AAW62706	Human brain expres	c 328	7	0.5	59	22	AAU50067	Propionibacterium
c 256	7	0.5	42	22	AAW75523	Human bone marrow	c 329	7	0.5	59	22	AAU30424	Novel human secret
c 257	7	0.5	42	22	AAW35630	Peptide #9667 enco	c 330	7	0.5	60	22	AAW67556	Propionibacterium
c 258	7	0.5	42	22	AAW65088	Gene #19 associate	c 331	7	0.5	60	22	AAW88937	Human immune/haema
c 259	7	0.5	42	23	ABG45074	Human peptide enco	c 332	7	0.5	60	22	AAW91442	Human immune/haema
c 260	7	0.5	44	20	AAW12028	Human 5' EST secre	c 333	7	0.5	61	22	AAW60201	Propionibacterium
c 261	7	0.5	44	22	ABG30028	Novel human diagno	c 334	7	0.5	61	22	AAU62142	Propionibacterium
c 262	7	0.5	46	22	ABG05181	Novel human diagno	c 335	7	0.5	61	22	AAU67180	Propionibacterium
c 263	7	0.5	47	22	ABW41064	Peptide #8570 enco	c 336	7	0.5	61	22	ABW42326	Peptide #9832 enco
c 264	7	0.5	47	22	ABW17487	Human nervous syst	c 337	7	0.5	61	22	ABW25817	Protein #7816 enco
c 265	7	0.5	47	22	AAW61922	Human brain expres	c 338	7	0.5	61	22	AAW63211	Human brain expres
c 266	7	0.5	47	22	AAW74724	Human bone marrow	c 339	7	0.5	61	22	AAW76025	Human bone marrow
c 267	7	0.5	47	22	AAW34841	Peptide #8878 enco	c 340	7	0.5	61	22	AAW20772	Peptide #7206 enco
c 268	7	0.5	48	20	AAW02741	Human secreted pro	c 341	7	0.5	61	22	AAW24067	Human EST encoded
c 269	7	0.5	48	22	ABW43462	Peptide #10968 enc	c 342	7	0.5	61	22	AAW36131	Peptide #10168 enc
c 270	7	0.5	48	22	ABW26428	Protein #8427 enco	c 343	7	0.5	61	22	AAW90811	Human shear stress
c 271	7	0.5	48	22	AAW64392	Human brain expres	c 344	7	0.5	61	23	ABG45423	Human peptide enco
c 272	7	0.5	48	22	AAW77212	Human bone marrow	c 345	7	0.5	62	22	AAU50072	Propionibacterium
c 273	7	0.5	48	22	AAW21145	Peptide #7579 enco	c 346	7	0.5	63	22	ABG01370	Novel human diagno
c 274	7	0.5	48	22	AAW37350	Peptide #11387 enc	c 347	7	0.5	63	23	ABW41270	Human ovarian anti
c 275	7	0.5	48	23	ABG46225	Human peptide enco	c 348	7	0.5	64	20	AAW88672	Secreted protein e
c 276	7	0.5	49	22	ABW10598	Human pancreatic c	c 349	7	0.5	64	21	AAW59891	Arabidopsis thalia
c 277	7	0.5	49	22	AAW92351	Human digestive sy	c 350	7	0.5	64	22	ABG10533	Novel human diagno
c 278	7	0.5	50	22	AAU44491	Propionibacterium	c 351	7	0.5	64	22	ABW50439	Human secreted pro
c 279	7	0.5	50	22	AAW00550	Human secreted pro	c 352	7	0.5	65	22	AAW92183	C glutamicum prote
c 280	7	0.5	50	22	AAW00968	Human bone marrow	c 353	7	0.5	66	21	AAW56342	Human secreted pro
c 281	7	0.5	50	22	ABW67768	Fragment from a wh	c 354	7	0.5	66	21	AAW32992	Pinus radiata tran
c 282	7	0.5	50	23	ABP07947	Human ORFX protein	c 355	7	0.5	66	21	AAW38527	Human secreted pro
c 283	7	0.5	51	21	AAW41803	Human ORFX ORF1567	c 356	7	0.5	66	22	AAW82678	Human immune/haema
c 284	7	0.5	51	22	AAU48303	Propionibacterium	c 357	7	0.5	66	22	AAW90726	C glutamicum prote
c 285	7	0.5	51	22	ABW42254	Peptide #9760 enco	c 358	7	0.5	67	22	ABG28031	Novel human diagno
c 286	7	0.5	51	22	AAW75954	Human bone marrow	c 359	7	0.5	68	21	AAW56982	Human prostate can
c 287	7	0.5	52	22	AAU58684	Propionibacterium	c 360	7	0.5	68	21	AAW19652	Arabidopsis thalia

361	7	0.5	68	22	AAU44247	Propionibacterium	c 434	7	0.5	91	22	AAAG75213	Human colon cancer
362	7	0.5	68	22	AAU53359	Propionibacterium	c 435	7	0.5	92	22	ABBI15424	Human nervous syst
c 363	7	0.5	68	22	AAM84261	Human immune/haema	436	7	0.5	93	21	ABA42262	Human ORFX ORF2026
364	7	0.5	68	22	AAM91282	Human immune/haema	437	7	0.5	93	22	ABB37831	Peptide #5337 enco
c 365	7	0.5	68	23	ABP34368	Human ORF3341 prot	438	7	0.5	93	22	ABB23097	Protein #5096 enco
366	7	0.5	68	23	ABP34381	Human ORF3354 prot	439	7	0.5	93	22	AAM58451	Human brain expres
c 367	7	0.5	69	22	AAU48937	Propionibacterium	440	7	0.5	93	22	AAM70943	Human bone marrow
c 368	7	0.5	69	22	AAU65779	Propionibacterium	441	7	0.5	93	22	AAM89523	Human immune/haema
369	7	0.5	69	23	ABP09342	Human ORFX protein	442	7	0.5	93	22	AAO03827	Human polypeptide
370	7	0.5	70	20	AAU13154	Human secreted pro	443	7	0.5	93	22	AAM31233	Peptide #5270 enco
c 371	7	0.5	71	21	AAG11076	Arabidopsis thalia	444	7	0.5	93	22	AAU21634	Novel human neopla
c 372	7	0.5	71	21	AAG56917	Arabidopsis thalia	445	7	0.5	93	23	ABG40742	Human peptide enco
c 373	7	0.5	71	22	AAM85033	Human immune/haema	446	7	0.5	93	23	ABP08264	Human ORFX protein
c 374	7	0.5	71	22	AAO13297	Human polypeptide	447	7	0.5	94	21	AAB58873	Breast and ovarian
375	7	0.5	72	22	AAM89498	Human immune/haema	448	7	0.5	94	21	AAG57637	Arabidopsis thalia
376	7	0.5	72	22	AAE03555	Human trefoil doma	449	7	0.5	94	22	AAU63724	Propionibacterium
c 377	7	0.5	73	21	AAG44945	Zea mays protein f	450	7	0.5	94	22	AAU77166	Corticotrophin rel
c 378	7	0.5	74	18	AAM10031	Protein encoded by	c 451	7	0.5	95	21	AAB41343	Human ORFX ORF1107
c 379	7	0.5	74	22	ABBI10140	Human cDNA SEQ ID	452	7	0.5	95	22	AAU61984	Propionibacterium
c 380	7	0.5	74	22	AAU20962	Human novel foetal	c 453	7	0.5	95	22	AAM82534	Human immune/haema
c 381	7	0.5	74	22	AAE09685	Human gene 3 encod	c 454	7	0.5	95	23	ABP10444	Human ORFX protein
382	7	0.5	75	21	AAG23802	Arabidopsis thalia	455	7	0.5	95	23	ABP10444	Human secreted pro
c 383	7	0.5	75	21	AAG02356	Human secreted pro	c 456	7	0.5	96	23	ABP41307	Human ovarian anti
384	7	0.5	75	23	ABP32432	Human ORF1405 prot	c 457	7	0.5	97	21	AAU16757	Bacteriophage Dp-1
c 385	7	0.5	75	23	ABP33979	Human ORF2952 prot	458	7	0.5	97	22	AAU48843	Propionibacterium
c 386	7	0.5	75	23	ABP00942	Human ORFX protein	459	7	0.5	97	22	AAU55127	Propionibacterium
387	7	0.5	76	22	ABBI7866	Human nervous syst	460	7	0.5	98	22	AAU67006	Propionibacterium
388	7	0.5	77	21	AAG00855	Human secreted pro	461	7	0.5	98	22	AAU55100	Propionibacterium
c 389	7	0.5	77	23	ABP35279	Human ORF4252 prot	462	7	0.5	98	22	ABBI5982	Human nervous syst
c 390	7	0.5	77	23	ABP01001	Human ORFX protein	463	7	0.5	98	22	ABBI7004	Human nervous syst
391	7	0.5	78	21	AAB54296	Human pancreatic c	464	7	0.5	98	22	AAG76522	Human colon cancer
c 392	7	0.5	78	22	AAM86714	Human immune/haema	c 465	7	0.5	98	22	AAG89223	Human secreted pro
c 393	7	0.5	78	22	AAO05173	Human polypeptide	c 466	7	0.5	98	22	AAU00284	Immunogenic epitop
c 394	7	0.5	78	22	AAU06055	Cone snail O-supe	c 467	7	0.5	99	22	AAM88579	Human immune/haema
395	7	0.5	78	23	ABP77048	Human protein sequ	468	7	0.5	100	20	AAU48251	Human prostate can
396	7	0.5	78	23	ABG54933	Human albumin fusi	c 469	7	0.5	100	22	AAU55373	Propionibacterium
397	7	0.5	79	22	AAU48480	Propionibacterium	470	7	0.5	100	22	AAG70808	S cerevisiae apopt
c 398	7	0.5	79	22	AAU54703	Propionibacterium	471	7	0.5	100	23	AAG78368	Human NMDA-R2B C-t
c 399	7	0.5	79	22	AAU56267	Propionibacterium	c 472	7	0.5	101	20	AAU24327	Rabies microti BM
400	7	0.5	79	22	AAU62175	Propionibacterium	c 473	7	0.5	101	21	ABJ30218	B. microti BMNI-6
c 401	7	0.5	80	21	AAG10149	Arabidopsis thalia	474	7	0.5	101	21	ABG10148	Arabidopsis thalia
c 402	7	0.5	80	22	ABG06551	Novel human diagno	c 475	7	0.5	101	23	ABB88963	Babesia microti an
c 403	7	0.5	80	23	ABP79278	Human prostate spe	476	7	0.5	102	22	AAU39200	Propionibacterium
404	7	0.5	81	16	AAG70158	Streptococcus pneu	477	7	0.5	102	22	AAM87171	Human immune/haema
c 405	7	0.5	81	20	AAU76537	Human ovarian tumo	478	7	0.5	102	22	AAO07991	Human polypeptide
c 406	7	0.5	81	22	AAU31912	Novel human secret	479	7	0.5	103	20	AAU39782	Interleukin-18 rec
407	7	0.5	81	22	AAM88234	Human immune/haema	480	7	0.5	103	20	AAU29191	Amino acid sequenc
c 408	7	0.5	81	22	AAM89768	Human immune/haema	c 481	7	0.5	103	20	AAU12429	Human 5' EST secre
c 409	7	0.5	81	22	AAM91747	Human immune/haema	482	7	0.5	103	21	AAG33671	Arabidopsis thalia
410	7	0.5	82	21	AAG00854	Human secreted pro	c 483	7	0.5	105	20	AAU24328	Babesia microti BM
c 411	7	0.5	83	21	AAG23801	Arabidopsis thalia	c 484	7	0.5	105	20	AAB30219	B. microti BMNI-6
c 412	7	0.5	83	22	AAU52126	Arabidopsis thalia	c 485	7	0.5	105	22	ABP71450	Drosophila melanog
413	7	0.5	83	22	AAG61154	Propionibacterium	c 486	7	0.5	105	23	ABB88964	Babesia microti an
414	7	0.5	84	21	AAG57122	Arabidopsis thalia	c 487	7	0.5	106	20	AAU12739	Human 5' EST secre
c 415	7	0.5	86	22	AAM86217	Human immune/haema	488	7	0.5	106	22	AAU63316	Propionibacterium
c 416	7	0.5	87	20	AAU02893	Fragment of human	489	7	0.5	106	22	ABBI7891	Human nervous syst
417	7	0.5	87	22	AAU66176	Propionibacterium	c 490	7	0.5	107	22	AAU59177	Propionibacterium
c 418	7	0.5	87	22	AAE01272	Human gene 7 encod	491	7	0.5	107	22	AAU62833	Propionibacterium
c 419	7	0.5	87	23	ABG63698	Human albumin fusi	492	7	0.5	107	22	AAE03402	Human gene 10 enco
c 420	7	0.5	88	21	AAG44944	Zea mays protein f	493	7	0.5	107	23	ABG63684	Human albumin fusi
421	7	0.5	88	22	AAU48338	Propionibacterium	c 494	7	0.5	108	22	AAM89714	Human immune/haema
c 422	7	0.5	88	22	AAU61588	Propionibacterium	c 495	7	0.5	109	18	AAW20446	H. pylori transpor
423	7	0.5	88	22	AAG75640	Human colon cancer	c 496	7	0.5	109	18	AAW24662	Novel human diagno
c 424	7	0.5	89	22	ABP69846	Drosophila melanog	c 497	7	0.5	109	22	ABG05912	Streptococcus pneu
c 425	7	0.5	89	22	AAU54691	Propionibacterium	c 498	7	0.5	110	19	AAW55113	Propionibacterium
426	7	0.5	89	22	AAM58314	Human brain expres	499	7	0.5	110	22	AAU54045	Human secreted pro
c 427	7	0.5	89	22	AAM70788	Human bone marrow	c 500	7	0.5	110	23	AAB90570	Human albumin fusi
c 428	7	0.5	89	22	AAM31087	Peptide #5124 enco	c 501	7	0.5	110	23	ABG65469	Human secreted pro
c 429	7	0.5	89	22	AAM06195	Peptide #4877 enco	c 502	7	0.5	110	23	ABP54607	S. pneumoniae SP06
c 430	7	0.5	90	21	AAG17387	Arabidopsis thalia	c 503	7	0.5	111	22	ABP54607	B. microti antigen
c 431	7	0.5	90	21	AAG44943	Zea mays protein f	c 504	7	0.5	111	22	AAU41399	Propionibacterium
c 432	7	0.5	90	22	ABBI5742	Human nervous syst	c 505	7	0.5	111	22	ABG13261	Novel human diagno
c 433	7	0.5	91	22	AAO06967	Human polypeptide	c 506	7	0.5	111	22	AAU14804	Novel bone marrow

c 507	7	0.5	111	23	ABP00417	Human ORFX protein	580	7	0.5	131	21	AAG12896	Arabidopsis thalia
c 508	7	0.5	112	22	ABB04681	B. microti antigen	c 581	7	0.5	131	22	ABB11734	Human GPCR homolog
c 509	7	0.5	112	22	AAU57069	Propionibacterium	c 582	7	0.5	131	22	AAU18534	Human cytoskeletal
c 510	7	0.5	112	22	AAE03551	Human trefoil doma	c 583	7	0.5	131	22	AAE76585	Corynebacterium gl
c 511	7	0.5	114	21	AGL11367	Arabidopsis thalia	c 584	7	0.5	131	22	AAU21809	Novel human neopla
c 512	7	0.5	115	20	AAI45268	Human secreted pro	c 585	7	0.5	132	22	ABG03513	Novel human diagno
c 513	7	0.5	115	20	AAW88609	Secreted protein e	c 586	7	0.5	132	22	AAQ01227	Human polypeptide
c 514	7	0.5	115	22	ABB50376	Human secreted pro	c 587	7	0.5	133	20	AAI30737	Amino acid sequenc
c 515	7	0.5	115	22	ABB17087	Human nervous syst	c 588	7	0.5	133	22	AAU60129	Propionibacterium
c 516	7	0.5	116	19	AAW54001	Anti-CD4 antibody	c 589	7	0.5	133	22	AAU64797	Propionibacterium
c 517	7	0.5	116	21	AGL16996	Arabidopsis thalia	c 590	7	0.5	134	8	AAI71175	First protein chal
c 518	7	0.5	116	23	AAE18340	Human B7-like prot	c 591	7	0.5	134	9	AAI80018	Sequence of the 18
c 519	7	0.5	117	22	ABG09163	Novel human diagno	c 592	7	0.5	134	21	AAE56272	Human secreted pro
c 520	7	0.5	117	23	ABG35320	Thrombopoietin ago	c 593	7	0.5	134	22	AAE68139	Amino acid sequenc
c 521	7	0.5	117	23	AAW47640	Human MPL #1. Hom	c 594	7	0.5	134	23	AAW51942	Human TGFbeta prot
c 522	7	0.5	119	22	ABB69173	Drosophila melanog	c 595	7	0.5	135	21	AAE33445	Human PRO1155 prot
c 523	7	0.5	119	22	ABB69462	Drosophila melanog	c 596	7	0.5	135	21	AAI66739	Membrane-bound pro
c 524	7	0.5	119	22	AAU51147	Propionibacterium	c 597	7	0.5	135	22	AAU29245	Human PRO polypept
c 525	7	0.5	119	22	AAE89199	Human secreted pro	c 598	7	0.5	135	22	AAQ01506	Human polypeptide
c 526	7	0.5	120	13	AAE27497	WN1 222-5 antibody	c 599	7	0.5	135	22	AAE92481	C glutamicum prote
c 527	7	0.5	120	20	AAW94275	Human eIF-4E-bindi	c 600	7	0.5	135	22	AAE65262	Human PRO1155 (UNQ
c 528	7	0.5	120	21	AAI96148	Human eIF-4E bindi	c 601	7	0.5	135	23	AAE95507	Human angio genesis
c 529	7	0.5	120	22	AAU44645	Propionibacterium	c 602	7	0.5	135	23	ABB84901	Human PRO1155 prot
c 530	7	0.5	120	22	ABB03155	Human musculoskele	c 603	7	0.5	135	23	AAE83645	Human angio genesis
c 531	7	0.5	121	19	AAW75228	Human secreted pro	c 604	7	0.5	136	21	AAI93275	Human PRO protein,
c 532	7	0.5	121	19	AAW75212	Human secreted pro	c 605	7	0.5	136	22	ABG07959	Amino acid sequenc
c 533	7	0.5	121	20	AAW97213	A human neurokin B	c 606	7	0.5	136	22	ABG12201	Novel human diagno
c 534	7	0.5	121	20	AAW74413	HPMBO91 protein se	c 607	7	0.5	136	23	ABG35323	Novel human diagno
c 535	7	0.5	121	22	ABG10328	Novel human diagno	c 608	7	0.5	137	19	AAW77671	Thrombopoietin ago
c 536	7	0.5	121	22	ABG82176	S. epidermidis ope	c 609	7	0.5	137	21	AAW41833	30S ribosomal prot
c 537	7	0.5	122	20	AAE82380	Human neurokinin B	c 610	7	0.5	137	21	AAI71471	Arabidopsis thalia
c 538	7	0.5	122	20	AAW96144	Human preprotachyk	c 611	7	0.5	137	22	AAE11878	Human prostaglandi
c 539	7	0.5	122	21	AGG07128	Arabidopsis thalia	c 612	7	0.5	139	17	AAE89134	Human neurokinin B
c 540	7	0.5	122	21	AAE58442	Arabidopsis thalia	c 613	7	0.5	139	17	AAW17676	Human early placen
c 541	7	0.5	122	21	AAE59389	Arabidopsis thalia	c 614	7	0.5	139	18	AAW69168	Human relaxin-rela
c 542	7	0.5	122	21	AAE661242	Arabidopsis thalia	c 615	7	0.5	139	20	AAI26926	Zinsi protein. Ho
c 543	7	0.5	122	21	AAE73201	Arabidopsis thalia	c 616	7	0.5	139	20	AAW99574	Human insulin-like
c 544	7	0.5	122	23	ABP38750	Inhibin balpha C-t	c 617	7	0.5	139	22	ABG05772	Human early placen
c 545	7	0.5	123	20	AAI39787	Staphylococcus epi	c 618	7	0.5	139	22	AAO04333	Novel human diagno
c 546	7	0.5	123	21	AAE53457	Interleukin-18 rec	c 619	7	0.5	140	21	AAE44909	Human polypeptide
c 547	7	0.5	123	21	AAE333846	Human colon cancer	c 620	7	0.5	140	22	AAE82356	zea mays protein f
c 548	7	0.5	123	21	AAE40493	Human secreted pro	c 621	7	0.5	140	23	AAW47842	S. epidermidis ope
c 549	7	0.5	123	21	AAE12897	Human ORFX ORF257	c 622	7	0.5	140	23	AAI27815	Human protein sequ
c 550	7	0.5	123	22	AAE90658	Arabidopsis thalia	c 623	7	0.5	141	20	AAI64529	Human secreted pro
c 551	7	0.5	123	23	ABP01509	Human lipocolon (L	c 624	7	0.5	141	22	AAO02557	Propionibacterium
c 552	7	0.5	124	21	AGG07127	Human ORFX protein	c 625	7	0.5	141	23	AAE68333	Human polypeptide
c 553	7	0.5	124	21	AAE57121	Arabidopsis thalia	c 626	7	0.5	143	23	ABF39916	Human lipocalin ho
c 554	7	0.5	124	21	AAE58441	Arabidopsis thalia	c 627	7	0.5	143	23	ABF39916	Staphylococcus epi
c 555	7	0.5	124	21	AAE59295	Arabidopsis thalia	c 628	7	0.5	145	21	AAG12895	Staphylococcus epi
c 556	7	0.5	124	21	AAE59388	Arabidopsis thalia	c 629	7	0.5	145	21	AAG21533	Arabidopsis thalia
c 557	7	0.5	124	21	AAE61241	Arabidopsis thalia	c 630	7	0.5	145	21	AAG25285	Arabidopsis thalia
c 558	7	0.5	124	22	AAE64677	Arabidopsis thalia	c 631	7	0.5	145	21	AAG29658	Arabidopsis thalia
c 559	7	0.5	125	21	AAE34558	Human secreted pro	c 632	7	0.5	145	22	ABE69035	Arabidopsis thalia
c 560	7	0.5	126	21	AAE44910	Arabidopsis thalia	c 633	7	0.5	146	21	AAG24496	Drosophila melanog
c 561	7	0.5	126	22	AAU50028	zea mays protein f	c 634	7	0.5	146	21	AAG36364	Arabidopsis thalia
c 562	7	0.5	126	22	AAI71848	Propionibacterium	c 635	7	0.5	146	21	AAG43677	Arabidopsis thalia
c 563	7	0.5	127	20	AAI24088	Human olfactory re	c 636	7	0.5	146	21	ABG15019	Arabidopsis thalia
c 564	7	0.5	127	22	AAU55399	Salmonella typhimu	c 637	7	0.5	146	22	ABG15019	Novel human diagno
c 565	7	0.5	127	22	ABG07966	Propionibacterium	c 638	7	0.5	146	22	AAE27696	Novel human diagno
c 566	7	0.5	128	21	AAE19639	Novel human diagno	c 639	7	0.5	146	22	AAE87060	Human immune/haema
c 567	7	0.5	128	21	AAE19643	Mouse monoclonal a	c 640	7	0.5	146	22	AAU16923	Human novel secret
c 568	7	0.5	128	21	AAE30201	Chimeric antibody	c 641	7	0.5	147	22	ABE52739	Human novel secret
c 569	7	0.5	128	22	ABE68852	B. microti BMNI-16	c 642	7	0.5	148	21	AAE23800	Escherichia coli p
c 570	7	0.5	128	22	AAU51285	Drosophila melanog	c 643	7	0.5	148	22	AAO06229	Arabidopsis thalia
c 571	7	0.5	128	22	ABG01132	Propionibacterium	c 644	7	0.5	149	21	AAE18616	Human polypeptide
c 572	7	0.5	128	23	ABB88947	Novel human diagno	c 645	7	0.5	149	21	AAG36363	Amino acid sequenc
c 573	7	0.5	129	20	AAI37136	Babesia microti an	c 646	7	0.5	149	22	AAE00855	Arabidopsis thalia
c 574	7	0.5	129	21	AAI33005	Amino acid sequenc	c 647	7	0.5	150	21	AAG82675	S. epidermidis ope
c 575	7	0.5	129	21	AAE34557	Arabidopsis thalia	c 648	7	0.5	150	22	AAG24495	Arabidopsis thalia
c 576	7	0.5	129	21	AAE00096	Arabidopsis thalia	c 649	7	0.5	150	21	AAE43676	Arabidopsis thalia
c 577	7	0.5	129	22	ABG10326	Human secreted pro	c 650	7	0.5	150	22	ABG01944	Novel human diagno
c 578	7	0.5	130	22	ABG06878	Novel human diagno	c 651	7	0.5	151	21	AAE42294	Human ORFX ORF2058
c 579	7	0.5	130	23	ABB55608	Novel human diagno	c 652	7	0.5	151	21	AAE18952	zea mays protein f
						Lactococcus lactis				151	23	ABP01033	Human ORFX protein

c 799	7	0.5	195	22	AAW79640	Human protein SEQ	c 872	7	0.5	222	21	AAAG07828	Arabidopsis thalia
c 800	7	0.5	195	22	AAW95863	Human protein sequ	c 873	7	0.5	222	21	AAAG16141	Arabidopsis thalia
c 801	7	0.5	195	22	AAW60776	Gene 47 related pe	c 874	7	0.5	222	21	AAAG27604	Arabidopsis thalia
c 802	7	0.5	195	23	ABW06799	Human ngPCR-Seq101	c 875	7	0.5	222	22	AAU31028	Novel human secret
c 803	7	0.5	196	22	AAU87444	Novel central nerv	c 876	7	0.5	222	22	AAU03629	Group B Streptococ
c 804	7	0.5	196	22	AAU35232	Enterococcus faeca	c 877	7	0.5	222	23	ABP26468	Streptococcus poly
c 805	7	0.5	196	22	AAU18648	Renal and cardiova	c 878	7	0.5	223	21	AAW57038	Human prostate can
c 806	7	0.5	197	22	ABG08070	Novel human diagno	c 879	7	0.5	223	21	AAW57038	Arabidopsis thalia
c 807	7	0.5	198	22	AAW63455	Propionibacterium	c 880	7	0.5	223	22	ABG23067	Novel human diagno
c 808	7	0.5	198	22	AAU63455	Propionibacterium	c 881	7	0.5	223	22	AAW40614	Human polypeptide
c 809	7	0.5	198	22	AAU69495	Human purified sec	c 882	7	0.5	224	20	AAW74086	Human prostate tum
c 810	7	0.5	198	22	AAU17112	Novel signal trans	c 883	7	0.5	224	22	ABG01822	Novel human diagno
c 811	7	0.5	198	23	AAW68329	Human lipocalin ho	c 884	7	0.5	225	22	ABG75290	Gene 3 human secre
c 812	7	0.5	200	20	AAW86634	Secreted protein e	c 885	7	0.5	225	13	ABP42795	Human ovarian anti
c 813	7	0.5	200	21	AAW53277	Human colon cancer	c 886	7	0.5	226	12	AAW13757	Prolactin. Mus mu
c 814	7	0.5	200	22	AAU33308	Staphylococcus aur	c 887	7	0.5	226	20	AAW39788	Interleukin-18 rec
c 815	7	0.5	200	22	AAU36985	Staphylococcus aur	c 888	7	0.5	226	21	AAW19331	Arabidopsis thalia
c 816	7	0.5	200	22	AAU37185	Staphylococcus aur	c 889	7	0.5	226	22	ABG03260	Novel human diagno
c 817	7	0.5	200	22	ABW50401	Human secreted pro	c 890	7	0.5	226	23	ABP06325	Human ORFX protein
c 818	7	0.5	200	23	AAO20048	Staphylococcus aur	c 891	7	0.5	227	20	AAW05312	S. aureus protein
c 819	7	0.5	200	23	AAO20069	Staphylococcus aur	c 892	7	0.5	227	22	ABW71814	Drosophila melanog
c 820	7	0.5	202	21	AAW96806	Streptococcus agal	c 893	7	0.5	227	22	ABG04921	Novel human diagno
c 821	7	0.5	202	22	AAU63393	Propionibacterium	c 894	7	0.5	228	23	ABF33032	Human ORF2005 prot
c 822	7	0.5	202	22	AAU66462	Propionibacterium	c 895	7	0.5	228	23	ABW5417	Lactococcus lactis
c 823	7	0.5	203	22	ABG12756	Novel human diagno	c 896	7	0.5	229	20	AAW27341	Group B Streptococ
c 824	7	0.5	204	21	AAW24726	Arabidopsis thalia	c 897	7	0.5	229	21	AAW31498	Arabidopsis thalia
c 825	7	0.5	204	23	ABW77327	Human growth hormo	c 898	7	0.5	230	22	AAW81587	S. epidermidis ope
c 826	7	0.5	205	21	AAW09892	Human glycine tran	c 899	7	0.5	231	12	AAW13518	M.ivanovii CORA.
c 827	7	0.5	205	21	AAW21164	Human glycine tran	c 900	7	0.5	231	21	AAW26310	Arabidopsis thalia
c 828	7	0.5	205	21	AAW21165	Human glycine tran	c 901	7	0.5	231	21	AAW37677	Arabidopsis thalia
c 829	7	0.5	205	21	AAW21166	Human glycine tran	c 902	7	0.5	231	21	AAW93280	Amino acid sequenc
c 830	7	0.5	205	21	AAW21167	Human glycine tran	c 903	7	0.5	231	21	AAW93300	Amino acid sequenc
c 831	7	0.5	205	21	AAW21168	Human glycine tran	c 904	7	0.5	231	21	AAW74392	Neisseria gonorrhoe
c 832	7	0.5	205	21	AAW21169	Human glycine tran	c 905	7	0.5	231	21	AAW74393	Neisseria meningit
c 833	7	0.5	205	22	ABW73956	Canine mutant RPE6	c 906	7	0.5	231	21	AAW74394	Neisseria meningit
c 834	7	0.5	206	21	ABW12119	Hydrophobic domain	c 907	7	0.5	231	22	AAW58594	Propionibacterium
c 835	7	0.5	206	22	ABW17052	Drosophila melanog	c 908	7	0.5	231	23	AAW20325	Human B7-H8 protei
c 836	7	0.5	207	20	AAW39784	Interleukin-18 rec	c 909	7	0.5	232	19	AAW75021	Human secreted pro
c 837	7	0.5	207	22	ABG29046	Novel human diagno	c 910	7	0.5	232	19	AAW75021	Human secreted pro
c 838	7	0.5	208	21	AAW16142	Arabidopsis thalia	c 911	7	0.5	233	21	AAW22269	ze mays protein f
c 839	7	0.5	208	21	AAW26312	Arabidopsis thalia	c 912	7	0.5	233	21	AAW09752	IGFBP-1 amino acid
c 840	7	0.5	208	21	AAW37679	Arabidopsis thalia	c 913	7	0.5	234	15	AAW62046	Human mature IGFBP
c 841	7	0.5	208	22	AAU18536	Human cytoskeletal	c 914	7	0.5	234	21	AAW09617	Insulin like growt
c 842	7	0.5	208	22	AAW23725	Human EST encoded	c 915	7	0.5	235	21	AAW41562	Arabidopsis thalia
c 843	7	0.5	208	22	AAW23729	Human EST encoded	c 916	7	0.5	236	21	AAW30331	Arabidopsis thalia
c 844	7	0.5	208	22	AAW21823	Novel human neopla	c 917	7	0.5	237	21	AAW57100	Human prostate can
c 845	7	0.5	208	23	ABF30274	Streptococcus poly	c 918	7	0.5	237	21	AAW07827	Arabidopsis thalia
c 846	7	0.5	209	21	AAW42974	Human ORFX ORF2738	c 919	7	0.5	237	21	AAW27603	Arabidopsis thalia
c 847	7	0.5	210	22	ABW70342	Drosophila melanog	c 920	7	0.5	237	21	AAW36022	Arabidopsis thalia
c 848	7	0.5	211	21	AAW19332	Arabidopsis thalia	c 921	7	0.5	238	20	AAW07464	ze mays protein f
c 849	7	0.5	211	22	AAU27475	Human G-Protein Co	c 922	7	0.5	238	20	AAW07022	Human TS10q23.3 ge
c 850	7	0.5	213	23	AAU97619	D. Immitis cystein	c 923	7	0.5	238	22	ABG05347	Breast cancer asso
c 851	7	0.5	215	23	ABG64839	Human albumin fusi	c 924	7	0.5	238	23	AAW50339	Novel human diagno
c 852	7	0.5	215	23	AAU96215	Human secreted pro	c 925	7	0.5	241	20	AAW50339	ETEC CS4 pilus Csa
c 853	7	0.5	215	23	AAU96235	Human secreted pro	c 926	7	0.5	241	23	AAW73982	Human prostate tum
c 854	7	0.5	216	23	ABP25952	Streptococcus poly	c 927	7	0.5	242	23	ABP40309	Cysteine protease
c 855	7	0.5	217	22	AAW75021	Human colon cancer	c 928	7	0.5	243	23	ABW47788	Staphylococcus epi
c 856	7	0.5	218	19	AAW70322	Secreted protein C	c 929	7	0.5	244	20	AAW06716	Listeria monocytog
c 857	7	0.5	219	22	ABW12045	Human channel-rela	c 930	7	0.5	245	20	AAW06715	Antibody 12B5 sing
c 858	7	0.5	219	22	AAW40605	Human polypeptide	c 931	7	0.5	245	22	AAW25480	Antibody 10D10 sin
c 859	7	0.5	219	23	ABG64782	Human albumin fusi	c 932	7	0.5	245	22	AAW25480	Human protein sequ
c 860	7	0.5	219	23	AAO17197	Human secreted pro	c 933	7	0.5	245	22	AAW25480	C glutamic prote
c 861	7	0.5	219	23	AAO17230	Human secreted pro	c 934	7	0.5	246	22	AAW76520	Corynebacterium gl
c 862	7	0.5	220	21	AAW26311	Arabidopsis thalia	c 935	7	0.5	246	21	AAW15540	Human immune syste
c 863	7	0.5	220	21	AAW37678	Arabidopsis thalia	c 936	7	0.5	247	21	AAW31658	Arabidopsis thalia
c 864	7	0.5	220	22	AAW02481	Arabidopsis thalia	c 937	7	0.5	249	18	AAW31220	HIV-TAT protein tr
c 865	7	0.5	220	22	AAW65001	Human secreted pro	c 938	7	0.5	249	19	AAW26460	Protein of the inv
c 866	7	0.5	220	23	ABP40228	Staphylococcus epi	c 939	7	0.5	249	19	AAW76166	Seq ID 57 from US5
c 867	7	0.5	220	23	AAU78292	Human Pregnane X R	c 940	7	0.5	251	21	AAW16369	Protein #2 used fo
c 868	7	0.5	221	21	AAW18928	A novel polypeptid	c 941	7	0.5	251	21	AAW45627	Arabidopsis thalia
c 869	7	0.5	221	22	AAU29267	Human PRO polypept	c 942	7	0.5	251	22	AAW34822	E. coli cellular p
c 870	7	0.5	221	22	AAW47451	SMAP-1. Homo sapi	c 943	7	0.5	251	22	AAW06774	Human polypeptide
c 871	7	0.5	221	22	AAW38828	Human polypeptide	c 944	7	0.5	251	23	AAW91449	Haemophilus influe


```
QY 181 CCAGATCTTTCCAGCTTTCTGAGCCCAAGAGAATTTAGACGAAAGTGTCAATTTGGCAAGA 240
Db 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
QY 241 CTGGCCATCAATACGACCTTTGGAGAGCGCAGAGTGAAGTCAAGCTAGAAACGACTT 300
Db 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
QY 301 TCTCCTGATCAGATGAACACTCACCTAAATTTAAGTTTTCAGCTAACTTCTGCCAGGAT 360
Db 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
QY 361 AACCTCTGAAAGTCCCAAGCTCTAAAGAAAGCCCGAGGAGCAAAAGGCCACAGTAT 420
Db 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
QY 421 TGTCTGTAACCCAGTCCAAAAGATATTTTAAATGAAGCTCCGACTTCATTTGAAGAG 480
Db 141 CysSerGluThrGlnSerLysLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
QY 481 CTATCTCTCCTTTTCAATCCCAACAGCTCCAAAAGATTTAGACCTCGTGCCTGCCAAAAC 540
Db 161 LeuSerSerLeuPheLysSerHisSerSerLysArgIleArgProArgAlaCysLysAsn 180
QY 541 CACAAGAGTAACTGGAATCTCAAAACAAAGTTATGCAGGAAACAGCTCCAGTTTCTCA 600
Db 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
QY 601 GATCTGTCAAGAAAGACGAGAAGATCTTCGTTCCTCCATCCCTATCCCTCGGGATACCAGG 660
Db 201 AspLeuSerGluArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
QY 661 GATAATGAAGTGAATCACGCCCTGGACACAGCAGGAGCCAGAGCGGTGAAGCGGACGAG 720
Db 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
QY 721 GCTGCCAGTGGAGCGGTGGTGAGACACTACACACAGGCTTCCCTCTCATCTCTGTAC 780
Db 241 AlaAlaSerGluAlaAlaGlyArgThrThrProGlySerSerProSerSerLeuTyr 260
QY 781 TATGAGAACCTCTGGGGCAACCTCCCGGTTCACATCAAAAGTTACGGACGACAGAGTT 840
Db 261 TyrGluGluProLeuGluGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
QY 841 CCAGAGGAAGTCCAGTACAGTGGATTCATAGTGTAGTAATCCACACCTCAAGTA 900
Db 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
QY 901 AGTGCTACTGTGAAGCAAGAGCTTGAATAATCCCCAGATATTCACATCGTCCAGGCA 960
Db 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
QY 961 GGAATCTGCACACTGACCTTGGGAGAGCTTTGAGAGACGACGACGCTATTC 1020
Db 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
QY 1021 TGTCTTCTTCTAACATCTATGGACAGATTCCGACTTCTGCTGAGATTTATATAGAGGG 1080
Db 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
QY 1081 GTTCTTCTTCTGACTCAGAGCGCCCTTAACAAGAGAGATGAATCAATCCAGAAG 1140
Db 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
QY 1141 CCAATAGAGTGTATCTCTCCACCTACCTCTGAGTCAATCTCCAGCAGTACCCCAA 1200
Db 381 ProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValProGln 400
QY 1201 GCCCAGCATTTGGTGGCCCAACCTCTGTGGCAACCATCCAGCAGTGTACAGCCCCACC 1260
Db 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnCysGlnSerProThr
```

```
QY 1261 AATTACTTCAGGAGTGTGATGAAACCTATCATTCAGCTCTGTGTTTACAAAGATG 1320
Db 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaAlaProValPheThrLysMet 440
QY 1321 CTCAAAATTTGTCAGCTTCTGAGGGTCAGCTGGTGTCTTTGAATCCAGAGTAAAGGA 1380
Db 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
QY 1381 GCTCCATCTCTAAGGTTGAGTGTATAGAGAAGGAGCTTTAATAGAGATTTCTCCAGAT 1440
Db 461 AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
QY 1441 TTTAGGATTTTACAGAAAACCTCGATCCATCGCAGAGCCA 1482
Db 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluPro 494
RESULT 2
ABBI0330
ID ABBI0330 standard; Protein; 264 AA.
XX
AC ABBI0330;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 638.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
muscular disorder; reproductive disorder; gastrointestinal disorder;
pulmonary disorder; cardiovascular disorder; renal disorder;
proliferative disorder; inflammation.
XX
OS Homo sapiens.
XX
PN WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179055P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216890P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
```

PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 03-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 08-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234977P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.

PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 11-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 03-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
DR N-PSDB; ABA06552.
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
PS Claim 11; SEQ ID NO: 638; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
SQ Sequence 264 AA;

Alignment Scores:
Pred. No.: 9.49e-232 Length: 264
Score: 261.00 Matches: 261
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.83% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x ABB10330 (1-264)

QY 3178 GAAAGACACAAAGAGCCCTACAGAACGCTTTTTCGACACACATTTTCTGACGGCTCCT 3237
|||||
Db 4 GluArgAspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaPro 23

QY 3238 GGGATATGCTAGCTCATGAGGGCGCCTCTGTGCGCTGGAGCTGTAAAGTGAGTGGTTTA 3297
|||||
Db 24 GlyAspMetValAlaHisGluGlyArgLeuLeuGluCysArgLeuAspCysLysValSerGlyLeu 43

QY 3298 CCGCCCCCGAGCTGACATGGCTACTCAATGGCCAACTGTGTACCATGCTCCAC 3357
|||||
Db 44 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 63

Db	141	ArgGlyArgSerArgValGlnGluArgAspLysGluProLeuGlnGluArgPhePheArg	160	XX	02-AUG-2001.
Oy	3217	CCACATTTCTCGCAGGCTCCCTGGGATATGGTAGCTCATGAGGGCGCTCTGTGGCGTG	3276	PD	
Db	161	ProHisPheLeuGlnAlaProGlyAspMetValAlaHisGluGlyArgLeuCysArgLeu	180	PF	17-JAN-2001; 2001WO-US01326.
Oy	3277	GACTGTAAGTGAGTGGTTTACCGCCCGGAGCTGACA-----TGGCTACTCAATGGCC	3331	XX	31-JAN-2000; 2000US-0179065.
Db	181	AspCysLysValSerGlyLeuProProPro-----AspMetThrTrpLeuLeuAsnGlyG	199	PR	04-FEB-2000; 2000US-0180628.
Oy	3332	AACCTGTGCTACCCAGATCGCTCCACAGATCGTGGTCAGGAGACCGGATCCACTCTC	3391	PR	24-FEB-2000; 2000US-0184664.
Db	199	InProValLeuProAspAlaSerHisLysMetLeuValArgGluThrGlyValHisSerL	219	PR	02-MAR-2000; 2000US-0186350.
Oy	3392	TGCTCATTCACCCACTCACTCAGCGCAGCAGGAGACCTATAAGTGCATCGCTACCAACA	3451	PR	16-MAR-2000; 2000US-0189874.
Db	219	euLeuIleAspProLeuThrGlnArgAspAlaGlyThrTyrLysCysIleAlaThrAsnL	239	PR	17-MAR-2000; 2000US-0190076.
Oy	3452	AAACGGCGCAGAATTCTTTTAGTCTGGAGCTCTCTGTAGTACCCAAAGAGGTGAAGAAAG	3511	PR	18-APR-2000; 2000US-0198123.
Db	239	ysThrGlyGlnAsnSerPheSerLeuGluLeuSerValValAlaLysGluValLysLysA	259	PR	19-MAY-2000; 2000US-0205515.
Oy	3512	CACCTGTGATCTCGAGAACTACAGAACTCGGCTGTTCCGAGAGGCCACCCCGTGAGAC	3571	PR	07-JUN-2000; 2000US-0209467.
Db	259	laProValIleLeuGluLysLeuGlnAsnCysGlyValProGluGlyHisProValArgL	279	PR	28-JUN-2000; 2000US-0214886.
Oy	3572	TGGAGTGGCGGTGATAGCATGCCCCCACCCTGTCTTCTACTGGGAAGAAACAAATGAGA	3631	PR	30-JUN-2000; 2000US-0215135.
Db	279	euGluCysArgValIleGlyMetProProProValPheTyrTrpLysLysAspAsnGluT	299	PR	07-JUL-2000; 2000US-0216647.
Oy	3632	CCATCCCTTGACCCAGAGAGAGGATCAGTATGCACCCAGCACACACAGGGTATGCCCTGCC	3691	PR	11-JUL-2000; 2000US-0217487.
Db	299	hrIleProCysThrArgGluArgIleSerMetHisGlnAspThrThrGlyTyrAlaCysL	319	PR	11-JUL-2000; 2000US-0217496.
Oy	3692	TTCTCATTCACCCAGCCCAAGAAACAGACGCTGGATGTGTACGTTGTCCACCAAGAATG	3751	PR	14-JUL-2000; 2000US-0218290.
Db	319	euLeuIleGlnProAlaLysLysSerAspAlaGlyTrpTyrThrLeuSerAlaLysAsnG	339	PR	26-JUL-2000; 2000US-0220964.
Oy	3752	AAGCGGCGATCGTGCAGCTGCAGCTGGATATATACGTCAGTGGCACCATCAGA	3811	PR	14-AUG-2000; 2000US-0225214.
Db	339	luAlaGlyIleValSerCysThrAlaArgLeuAspIleTyrAlaGlnTrpHisHisGlnI	359	PR	14-AUG-2000; 2000US-0225266.
Oy	3812	TCCCACCCCGCATGCTCTCCGGCCCACTGGCAGTCGCTACGGATCTCTCACCAGTAAAG	3871	PR	14-AUG-2000; 2000US-0225287.
Db	359	leProProProMetSerValArgProSerGlySerArgTyrGlySerLeuThrSerLysG	379	PR	14-AUG-2000; 2000US-0225287.
Oy	3872	GACTTGACATATTTCTCCCTTTTCCCTCCATGGAAAGCACCATGGTGTATTCATGCTCTT	3931	PR	14-AUG-2000; 2000US-0225287.
Db	379	lyLeuAspIlePheSerAlaPheSerMetGluSerThrMetValTyrSerCysSers	399	PR	14-AUG-2000; 2000US-0225287.
Oy	3932	CTCGAGTGTAGTGGAGAGTGATGAATT	3960	PR	14-AUG-2000; 2000US-0225287.
Db	399	erArgSerValValGluSerAspGluLeu	408	PR	14-AUG-2000; 2000US-0225287.
RESULT 4					
AAU17989					
ID	AAU17989	standard; Protein; 261 AA.			
XX					
AC	AAU17989;				
XX					
DT	07-NOV-2001	(first entry)			
XX					
DE		Human immunoglobulin polypeptide SEQ ID No 134.			
XX					
KW		Immunoglobulin; signal transduction pathway protein; cancer;			
KW		antisense therapy; gene therapy; neurological disorder; renal disorder;			
KW		cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;			
KW		reproductive disorder; immune system disorder; proliferative disorder;			
XX		muscular disorder.			
OS		Homo sapiens.			
XX					
PN	WO200155315-A2.				

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX
XX

DR WPI; 2001-457725/49.
DR N-PSDB; AAS28777.
XX Isolated novel immunoglobulin polypeptide for monitoring the presence
PT and progression of diseases and for diagnosis -
XX
XX Claim 11; SEQ ID No 134; 551pp; English.
XX
CC Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the
CC invention. The polypeptides and their associated polynucleotides can be
CC used to diagnose a pathological condition or a susceptibility to a
CC pathological condition in a subject by determining the presence or
CC absence of a mutation in a DNA sequence or determining the presence or
CC amount of expression of the protein. Alternatively the identification of
CC a binding partner to a sequence allows determination of changes in
CC protein activity. The sequences can be used as research tools for
CC receptors or other signal transduction pathway proteins that interact
CC with the polypeptides of the invention and can be used to treat, prevent
CC or diagnose various types of disorders such as neurological disorders,
CC cardiovascular disorders, gastrointestinal disorders, reproductive
CC disorders, immune system disorders, renal disorders, muscular disorders,
CC pulmonary disorders, proliferative disorders and cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 261 AA;

Alignment Scores:
Pred. No.: 4.73e-156 Length: 261
Score: 179.00 Matches: 179
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.60% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x AAU17989 (1-261)

QY 3178 GAAAGAGACAAAGAGCCCTACAGGACGCTTTTCCGACACATTTCTCGAGGCTCCT 3237
Db 1 GluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPro 20

QY 3238 GGGATATGCTAGCTCATGAGGGCGCTCTGTCGGTGGACTGTAAGGTGAGTGTTTA 3297
Db 21 GlyAspMetValAlaHisGluGlyArgLeuGluCysArgLeuAspCysLysValSerGlyLeu 40

QY 3298 CCGCCCCCGAGTGACATGGCTACTCAATGGCCAACTGTGTACAGATGCTCCAC 3357
Db 41 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 60

QY 3358 AAGATGCTGTCAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCCACTCAGCGC 3417
Db 61 LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuLeuLeuLeuGlnArg 80

QY 3418 GAGCGAGGACCTATAGTGCATCGCTACCAACAAACCGGAGAGATTTCTTTAGTCTG 3477
Db 81 AspAlaGlyThrTyrlsCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 100

QY 3478 GAGCTCTCTAGTAGCCAAAGAGGTGAAGAAAGACCTGATCTGTGGAAGAACTACAG 3537
Db 101 GluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGln 120

QY 3538 AACTGCGGTGTTCCCGAAGGCCCCCGTGGAGTGGAGTGGCCGCGATAGGATGCC 3597
Db 121 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 140

QY 3598 CCACCTGTGTTCTACTGGAAGAAAGACAATGAGACCATCCCTTCCACAGAGAGAGGATC 3657
Db 141 ProProValPheTyrlsLysLysAspAsnGluThrIleProCysThrArgGluArgIle 160

QY 3658 AGTATGCACGAGGACACACAGGATGCTGCTGCTTCTCATTTACGAGCAAGAAA- 3714
Db 161 SerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLys 179

RESULT 5
ABBI0509
ID ABB10509 standard; Protein; 159 AA.
XX
AC ABB10509;
XX
DT 10-JAN-2002 (first entry)
XX
DE Human cDNA SEQ ID NO: 817.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
XX
OS Homo sapiens.
XX
PN W0200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225266P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228944P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234977P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249246P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.

PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 03-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-476161/51.
XX N-PSDB; ABA06731.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition -
XX
XX Claim 11; SEQ ID NO: 817; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
XX SQ Sequence 159 AA;

Alignment Scores:
Pred. No.: 4,31e-95 Length: 159
Score: 113.00 Matches: 113
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.59% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x ABB10509 (1-159)

QY 3424 GGGACCTATAGTCATCGCTACCAACAAACCGGGCAGAAATCTTTTAGTCTGGAGCTC 3483
Db 8 GlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeu 27

QY 3484 TCTGTACTAGCCAAAGAGGTGAAGAAACACCTGTGTATCTCTGGAACTACAGAACTGC 3543
Db 28 SerValValAlaLysGluValLysAlaProValIleLeuGluLysLeuGlnAsnCys 47

QY 3544 GGTGTTCCCGAAGCCACCCCGTGGAGTGGAGTGGCCGCTGATAGCATGCCCCACCT 3603
Db 48 GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProPro 67

QY 3604 GTGTTCTACTGGAGAAAGAACATGAGACCATCCCTTGCACGAGAGAGATCAGTATG 3663
Db 68 ValPheTyrTrpLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMet 87

QY 3664 CACGAGCACACAGAGGTATGCTGCTTCTCATTCAGCCAGCCAGCAAGAAATCAGAGCT 3723
Db 88 HisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAla 107

QY 3724 GGATGGTACACGTTGTGTCAGCAAGAAATGAAGCCGGCATC 3762
Db 108 GlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIle 120

RESULT 6
AAU18053
ID AAU18053 standard; Protein; 159 AA.

XX AAU18053;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immunoglobulin polypeptide SEQ ID NO 198.
XX
XX Immunoglobulin; signal transduction pathway protein; cancer;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
XX Homo sapiens.
XX
XX WO200155315-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01326.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 03-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234907.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0951030.
PR 05-DEC-2000; 2000US-0951988.
PR 05-DEC-2000; 2000US-0956719.
PR 06-DEC-2000; 2000US-0951479.
PR 08-DEC-2000; 2000US-0951856.
PR 08-DEC-2000; 2000US-0951868.
PR 08-DEC-2000; 2000US-0951869.
PR 08-DEC-2000; 2000US-0951989.
PR 08-DEC-2000; 2000US-0951990.
PR 11-DEC-2000; 2000US-0954097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457725/49.
XX N-PSDB; AAS28841.
XX Isolated novel immunoglobulin polypeptide for monitoring the presence
XX and progression of diseases and for diagnosis -
PS Claim 11; SEQ ID No 198; 551pp; English.
XX Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the
XX invention. The polypeptides and their associated polynucleotides can be
XX used to diagnose a pathological condition or a susceptibility to a
XX pathological condition in a subject by determining the presence or
XX absence of a mutation in a DNA sequence or determining the presence or
XX amount of expression of the protein. Alternatively the identification of
XX a binding partner to a sequence allows determination of changes in
XX protein activity. The sequences can be used as research tools for
XX receptors or other signal transduction pathway proteins that interact
XX with the polypeptides of the invention and can be used to treat, prevent
XX or diagnose various types of disorders such as neurological disorders,
XX cardiovascular disorders, gastrointestinal disorders, reproductive
XX disorders, immune system disorders, renal disorders, muscular disorders,
XX pulmonary disorders, proliferative disorders and cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 159 AA;

Alignment Scores:
Pred No.: 4,31e-95 Length: 159
Score: 113.00 Matches: 113
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.59% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x AAU18053 (1-159)
Qy 3424 GGGACCTATAAGTCATCGCTACCAACAAAACCGGCGAGAAATCTTTTAGTCTGGAGCTC 3483
|||||
Db 8 GlyThrTyrlLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeu 27

Qy 3484 TCTGTAGTAGCCAAAGAGGTGAAGAACACCTGTGATCCTGGAGAACTACAAACTGC 3543
|||||
Db 28 SerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCys 47

Qy 3544 GGTTCTCCGAAGCCACCCCGTGGAGTGGCGGTGATAGGCATGCCCCACCT 3603
|||||
Db 48 GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProProPro 67

Qy 3604 GTGTTCTACTGGAAGAAGACAATGAGACCATCCCTTCACCCAGAGAGAGATCAGTATG 3663
|||||
Db 68 ValPheTyrrTrrLysLysAspAsnGluThrIleProCysThrArgGluArgIleSerMet 87

Qy 3664 CACCAGGACACACAGGGTATGCCCTTCTCATTTCAGCCAGCCAGCAAGAAATCAGACGCT 3723
|||||
Db 88 HisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAla 107

QY 3724 GGATGTCACACGTGTGTCAGCCAGAAATGAAGCGGCATC 3762
|||||
Db 108 GlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIle 120

RESULT 7
AAB40875
ID AAB40875 standard; Protein; 130 AA.
XX AAB40875;
AC AAB40875;
XX AAB40875;
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF639 polypeptide sequence SEQ ID NO:1278.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipapillary; antipapillary; antipapillary; antipapillary;
KW anticonvulsant; osteopathic; antipapillary; antipapillary; antipapillary;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antidiabetic;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
OS Homo sapiens.
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX Shinkets RA, Leach M;
PI Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC75084.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 1115-1116; 5507pp; English.
PS
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipapillary; antipapillary; antipapillary; antipapillary;
CC osteopathic; anticonvulsant; antipapillary; antipapillary; antipapillary;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 130 AA;
Alignment Scores:
Pred. No.: 1,54e-65 Length: 130
Score: 81.00 Matches: 81
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.16% Indels: 0
DB: 21 Gaps: 0

US-09-818-990b-1 (1-3963) x AAB40875 (1-130)
QY 2542 CTGCTAGAGTGCACCATCCATCCATCCAGGATAGCGAAGAAAATACAAAGTCT 2601
|||||
Db 43 LeuProArgSerAlaProSerMetProSerGlnGlyLeuAlaLysLysAsnThrLysSer 62
2602 CCTCACCCAGTGAATGATGATTAACATTCGTGAACCTAAGAACCGAGTATTCGAGACTTG 2661
|||||
Db 63 ProGlnProValAsnAspAspAsnIleArgGluThrLysAsnAlaValIleArgAspLeu 82
2662 GGGAAAAAATAACTTTCAGTGTGTCAGACCAACACAGCAGGAGTACAAAATTCAGC 2721
|||||
Db 83 GlyLysLysIleThrPheSerAspValArgProAsnGlnGlnGluThrLysIleSerSer 102
2722 TTTGAGCAGAGGCTGATGATGAATGAAATAGAGTTTCGGCTTGAAGCTACTCTCTGTTGATGAA 2781
|||||
Db 103 PheGluGlnArgLeuMetAsnGluIleGluPheArgLeuGluArgThrProValAspGlu 122
QY 2782 TCA 2784
|||
Db 123 Ser 123

RESULT 8
ABP09582
ID ABP09582 standard; Protein; 50 AA.
XX
AC ABP09582;
XX
DT 24-JUN-2002 (first entry)
XX Human ORFX protein sequence SEQ ID NO:19146.
DE
DE Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
XX Homo sapiens.
XX
XX WO200192523-A2.
PN
XX
XX 06-DEC-2001.
PD
XX
XX 29-MAY-2001; 2001WO-US10836.
PF
XX
XX 30-MAY-2000; 2000US-206132P.
PR
XX 29-AUG-2000; 2000US-228716P.
PR
XX (CURA-) CURAGEN CORP.
PA
XX Shinkets RA, Leach MD;
PI Shinkets RA, Leach MD;
XX WPI; 2002-106308/14.
DR N-PSDB; ABN25334.
DR

XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
XX Disclosure; SEQ ID 19146; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ASN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 50 AA;

Alignment Scores:
Pred. No.: 3,66e-21 Length: 50
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.51% Indels: 0
DB: 23 Gaps: 0

US-09-818-990b-1 (1-3963) x ABP09582 (1-50)

QY 2122 AAGCAGGTGAAGCTCCTCATCACAGACGTTGCGTGGCGGAGATTTCTTC 2181
Db 17 LysGlnValLysAlaProSerSerGlnThrPheSerLeuAlaArgProLysTyrPhePhe 36

QY 2182 CCTCCAGAACACCGCCGACCACTGTGGCCCTTCC 2220
Db 37 ProSerThrAsnThrThrAlaAlaThrValAlaProSer 49

RESULT 9
AAB42762
ID AAB42762 standard; Protein: 433 AA.
XX
XX AAB42762;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF2526 polypeptide sequence SEQ ID NO:5052.
XX
XX Human; open reading frame; ORFX; detection: cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC76971.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4239-4240; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 433 AA;

Alignment Scores:
Pred. No.: 2,19e-08 Length: 433
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
-Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 21 Gaps: 0

US-09-818-990b-1 (1-3963) x AAB42762 (1-433)

QY 3034 GATGACGATGGCAACATACACCATCATGGAGCCACCCCGGGGAGAAATCAGCTGT 3090
Db 118 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 136

RESULT 10
AAB53464
ID AAB53464 standard; Protein: 544 AA.
XX

AC AAB53464;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen protein sequence SEQ ID NO:1004.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
OS Homo sapiens.
XX
XX WO20005351-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05883.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587534/55.
XX
XX N-PSDB; AAC98221.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
XX antigens, useful for the treatment, prevention, and diagnosis of colon
XX disorders such as colon cancer -
XX
XX Claim 11; Page 1585-1587; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX vulnary, nephrotropic, antiinfective and antibacterial activities, and
XX can be used in gene therapy. The colon cancer antigen polynucleotides,
XX proteins and antibodies to the proteins are useful for the prevention,
XX treatment and diagnosis of colon disorders, such as colon cancer. The
XX polynucleotides may be used in diagnostics and research, such as for
XX chromosome identification, and as hybridisation probes. The proteins
XX may also be used to prevent diseases such as neural disorders, immune
XX system disorders, muscular disorders, reproductive disorders,
XX gastrointestinal disorders, wounds, renal disorders, infectious
XX diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
XX AAB54007 represent sequences used in the exemplification of the present
XX invention.
XX
SQ Sequence 544 AA;
Alignment Scores:
Pred. No.: 2.12e-08 Length: 544
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 21 Gaps: 0
US-09-818-990B-1 (1-3963) x AAB53464 (1-544)
QY 3034 GATGACGATGCACTACACATCATGGCCACCAACCCAGGGGAGAGATCAGCTGT 3090
|||||
Db 229 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 247
RESULT 11
AAU01767
ID AAU01767 standard; Protein; 43 AA.

XX AAU01767;
AC
XX 18-JUL-2001 (first entry)
DT
XX
DE Human secreted protein #46.
XX
KW Human; secreted protein; immunogen; antibody; diagnosis;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW fungal infection; corneal infection; wound healing; cell culture;
KW epithelial cell proliferation; skin ageing; transplantation;
KW tissue regeneration; chemotaxis; food additive.
XX
XX Homo sapiens.
XX
XX WO200123546-A1.
XX
XX 05-APR-2001.
XX
XX 26-SEP-2000; 2000WO-US26323.
XX
XX 27-SEP-1999; 99US-0155805.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben S, Komatsoulis GA;
XX
XX WPI; 2001-266150/27.
XX
XX N-PSDB; AAS02397.
XX
XX Nucleic acids encoding 37 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX
XX Disclosure; Page 457; 494pp; English.
XX
XX The sequence represents a human secreted protein of the invention. The
XX polynucleotides, polypeptides and antibodies raised against them are used
XX to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. The
XX polynucleotides and antibodies are also used in diagnosing a pathological
XX condition or susceptibility to a pathological condition. The antibodies
XX can also be used in alleviating symptoms associated with the disorders
XX and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbent assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angiogenesis, nervous system disorders e.g. Alzheimer's disease,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities.
XX
SQ Sequence 43 AA;
Alignment Scores:
Pred. No.: 1.85e-05 Length: 43
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.22% Indels: 0
DB: 22 Gaps: 0
US-09-818-990B-1 (1-3963) x AAU01767 (1-43)
QY 3739 TCAGCCAAGATGAAGCGGCATCGTGTGTCGACTGCCAGCTGGAT 3786
|||||

Db 22 SerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 37

RESULT 12

AAU01765
ID AAU01765 standard; Protein; 44 AA.

AC AAU01765;

DT 18-JUL-2001 (first entry)

DE Human secreted protein #44.

KW Human; secreted protein; immunogen; antibody; diagnosis;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW cardiovascular disorder; cerebrovascular disorder; cerebral
KW angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW fungal infection; corneal infection; wound healing; cell culture;
KW epithelial cell proliferation; skin ageing; transplantation;
KW tissue regeneration; chemotaxis; food additive.

OS Homo sapiens.

XX WO200123546-A1.

XX 05-APR-2001.

XX 26-SEP-2000; 2000WO-US26323.

XX 27-SEP-1999; 99US-0155805.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben S, Komatsoulis GA;

XX WPI; 2001-266150/27.

DR N-PSDB; AAS02397.

XX Nucleic acids encoding 37 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -

PS Disclosure; Page 456-457; 494pp; English.

XX The sequence represents a human secreted protein of the invention. The
CC polynucleotides, polypeptides and antibodies raised against them are used
CC to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC polynucleotides and antibodies are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. The antibodies
CC can also be used in alleviating symptoms associated with the disorders
CC and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.

XX Sequence 44 AA;

Alignment Scores:

Pred. No.: 1.84e-05 Length: 44
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.22% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x AAU01765 (1-44)

QY 3739 TCACCAAGATGAAGCGGCATCGTGTGCACTGCCAGGCTGGAT 3786

DB 28 SerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 43
|||||

RESULT 13

AAU01763

ID AAU01763 standard; Protein; 49 AA.

XX AAU01763;

XX 18-JUL-2001 (first entry)

DE Human secreted protein #42.

KW Human; secreted protein; immunogen; antibody; diagnosis;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW cardiovascular disorder; cerebrovascular disorder; cerebral
KW angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW fungal infection; corneal infection; wound healing; cell culture;
KW epithelial cell proliferation; skin ageing; transplantation;
KW tissue regeneration; chemotaxis; food additive.

OS Homo sapiens.

XX WO200123546-A1.

XX 05-APR-2001.

XX 26-SEP-2000; 2000WO-US26323.

XX 27-SEP-1999; 99US-0155805.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben S, Komatsoulis GA;

XX WPI; 2001-266150/27.

DR N-PSDB; AAS02397.

XX Nucleic acids encoding 37 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -

PS Disclosure; Page 456; 494pp; English.

XX The sequence represents a human secreted protein of the invention. The
CC polynucleotides, polypeptides and antibodies raised against them are used
CC to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC polynucleotides and antibodies are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. The antibodies
CC can also be used in alleviating symptoms associated with the disorders
CC and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.

XX Sequence 49 AA;

Alignment Scores:

Pred. No.: 1.81e-05 Length: 49

Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1-22% Indels: 0
DB: 22 Gaps: 0

US-09-818-990B-1 (1-3963) x AAU01763 (1-49)

QY 3739 TCAGCCAGGAATGAAGCGGCATCGTCTGTCGACATGCCAGCTGGAT 3786

|||||
Db 28 SerAlaLysAsnGluAlaGlyIleValSerCysThrAlaArgLeuAsp 43

RESULT 14

AAU06795

ID AAU06795 standard; Protein; 104 AA.

XX

AC AAU06795;

XX

XX 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 1003.

XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW neutrotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation.

XX Homo sapiens.

OS

XX WO20015339-A2.

PN

XX 02-AUG-2001.

PD

XX 25-JAN-2001; 2001WO-US02723.

PF

XX 25-JAN-2000; 2000US-0491404.

PR 15-SEP-2000; 2000US-0563870.

PR 06-NOV-2000; 2000US-0707351.

XX (HYSE-) HYSEQ INC.

PA

XX Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;

PI Liu C, Asundi V, Zhou P, Werhman T;

PI

XX WPI: 2001-465571/50.

DR N-PSDB; AAH94470.

XX

PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -

XX Example 4; Page 577-578; 715pp; English.

XX The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence is a polypeptide encoded by a cDNA assembled using
CC an expressed sequence tag (EST) found to be expressed in human
CC foetal tissue cDNA libraries.

XX SQ Sequence 104 AA;

XX

Alignment Scores:

Pred. No.: 0.00113 Length: 104
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.06% Indels: 0

DB: 22 Gaps: 0
US-09-818-990B-1 (1-3963) x AAU06795 (1-104)

QY 3034 GATGACAGTGGCACTACACCATCATGCGAGCCAAACCCCCAG 3075
|||||
Db 73 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGln 86

RESULT 15

ABP07380

ID ABP07380 standard; Protein; 162 AA.

XX

AC ABP07380;

XX

XX 25-JUN-2002 (first entry)

DE

XX Human ORFX protein sequence SEQ ID NO:14742.

XX

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.

XX Homo sapiens.

OS

XX WO200192523-A2.

PN

XX 06-DEC-2001.

PD

XX 29-MAY-2001; 2001WO-US10836.

PF

XX 30-MAY-2000; 2000US-206132P.

PR 29-AUG-2000; 2000US-228716P.

PR

XX (CURA-) CURAGEN CORP.

XX

XX Shimkets RA, Leach MD;

PI

XX WPI; 2002-106308/14.

DR

XX N-PSDB; ABN23132.

DR

XX

XX Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,

PT hyperproliferative disorders and autoimmune disorders -

PT

XX Disclosure; SEQ ID 14742; 1037pp; English.

XX

CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is I-I1491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.

CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 162 AA;

Alignment Scores:
Pred. NO.: 0.00106 Length: 162
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.06% Indels: 0
DB: 23 Gaps: 0

US-09-818-990B-1 (1-3963) x ABP07380 (1-162)

OY 1429 GATTCTCCAGATTTTAGGATTTTACAGAAAAACCTCGATCC 1470
Db 127 AspSerProAspPheArgIleLeuGlnLysLysProArgSer 140

Search completed: November 30, 2002, 18:43:24
Job time : 103 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:41:29 ; Search time 23 Seconds
(without alignments)
10139.395 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atycaagacgacagcataga.....tggagagtgtgaacttta 3963

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/Cgn2_1/USPTO.spool/US09818990/runat_26112002_093223_22098/app_query_fasta_1.4103
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=oligo.raii -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09818990.ecgn_1.1.20 -runat_26112002_093223_22098 -NCPU=6
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	8	0.6	192	4	US-09-149-476-477
C 2	8	0.6	305	4	US-09-500-569-2
C 3	8	0.6	352	2	US-08-440-845D-5
C 4	8	0.6	352	4	US-08-868-458-5
C 5	8	0.6	352	4	US-09-134-262-2
C 6	8	0.6	352	4	US-09-359-070-2
C 7	8	0.6	362	1	US-08-464-523B-32
C 8	8	0.6	362	2	US-08-948-176-25
C 9	8	0.6	362	2	US-08-440-845D-6
C 10	8	0.6	362	4	US-08-868-458-6
C 11	8	0.6	362	4	US-09-303-592-2
C 12	8	0.6	362	4	US-09-303-592-4
C 13	8	0.6	362	4	US-09-303-592-4
C 14	8	0.6	362	4	US-09-303-592-4
C 15	8	0.6	362	4	US-09-303-592-4
C 16	8	0.6	362	4	US-09-303-592-4
C 17	8	0.6	362	4	US-09-303-592-4
C 18	8	0.6	362	4	US-09-303-592-4
C 19	8	0.6	362	4	US-09-303-592-4
C 20	8	0.6	362	4	US-09-303-592-4
C 21	8	0.6	362	4	US-09-303-592-4
C 22	8	0.6	362	4	US-09-303-592-4
C 23	8	0.6	362	4	US-09-303-592-4
C 24	8	0.6	362	4	US-09-303-592-4
C 25	8	0.6	362	4	US-09-303-592-4
C 26	8	0.6	362	4	US-09-303-592-4
C 27	8	0.6	362	4	US-09-303-592-4
C 28	8	0.6	362	4	US-09-303-592-4
C 29	8	0.6	362	4	US-09-303-592-4
C 30	8	0.6	362	4	US-09-303-592-4
C 31	8	0.6	362	4	US-09-303-592-4
C 32	8	0.6	362	4	US-09-303-592-4
C 33	8	0.6	362	4	US-09-303-592-4
C 34	8	0.6	362	4	US-09-303-592-4
C 35	8	0.6	362	4	US-09-303-592-4
C 36	8	0.6	362	4	US-09-303-592-4
C 37	8	0.6	362	4	US-09-303-592-4
C 38	8	0.6	362	4	US-09-303-592-4
C 39	8	0.6	362	4	US-09-303-592-4
C 40	8	0.6	362	4	US-09-303-592-4
C 41	8	0.6	362	4	US-09-303-592-4
C 42	8	0.6	362	4	US-09-303-592-4
C 43	8	0.6	362	4	US-09-303-592-4
C 44	8	0.6	362	4	US-09-303-592-4
C 45	8	0.6	362	4	US-09-303-592-4
C 46	8	0.6	362	4	US-09-303-592-4
C 47	8	0.6	362	4	US-09-303-592-4
C 48	8	0.6	362	4	US-09-303-592-4
C 49	8	0.6	362	4	US-09-303-592-4
C 50	8	0.6	362	4	US-09-303-592-4
C 51	8	0.6	362	4	US-09-303-592-4
C 52	8	0.6	362	4	US-09-303-592-4
C 53	8	0.6	362	4	US-09-303-592-4
C 54	8	0.6	362	4	US-09-303-592-4
C 55	8	0.6	362	4	US-09-303-592-4
C 56	8	0.6	362	4	US-09-303-592-4
C 57	8	0.6	362	4	US-09-303-592-4
C 58	8	0.6	362	4	US-09-303-592-4
C 59	8	0.6	362	4	US-09-303-592-4
C 60	8	0.6	362	4	US-09-303-592-4
C 61	8	0.6	362	4	US-09-303-592-4
C 62	8	0.6	362	4	US-09-303-592-4
C 63	8	0.6	362	4	US-09-303-592-4
C 64	8	0.6	362	4	US-09-303-592-4
C 65	8	0.6	362	4	US-09-303-592-4
C 66	8	0.6	362	4	US-09-303-592-4
C 67	8	0.6	362	4	US-09-303-592-4
C 68	8	0.6	362	4	US-09-303-592-4
C 69	8	0.6	362	4	US-09-303-592-4
C 70	8	0.6	362	4	US-09-303-592-4
C 71	8	0.6	362	4	US-09-303-592-4
C 72	8	0.6	362	4	US-09-303-592-4
C 73	8	0.6	362	4	US-09-303-592-4
C 74	8	0.6	362	4	US-09-303-592-4
C 75	8	0.6	362	4	US-09-303-592-4
C 76	8	0.6	362	4	US-09-303-592-4
C 77	8	0.6	362	4	US-09-303-592-4
C 78	8	0.6	362	4	US-09-303-592-4
C 79	8	0.6	362	4	US-09-303-592-4
C 80	8	0.6	362	4	US-09-303-592-4
C 81	8	0.6	362	4	US-09-303-592-4
C 82	8	0.6	362	4	US-09-303-592-4
C 83	8	0.6	362	4	US-09-303-592-4
C 84	8	0.6	362	4	US-09-303-592-4
C 85	8	0.6	362	4	US-09-303-592-4

Sequence 54, Appl
Sequence 13, Appl
Sequence 53, Appl
Sequence 52, Appl
Sequence 2, Appl
Sequence 7, Appl
Sequence 4425, Ap
Sequence 1, Appl
Sequence 15, Appl
Sequence 10, Appl
Sequence 2, Appl
Sequence 35, Appl
Sequence 521, App
Sequence 378, App
Sequence 1, Appl
Sequence 1, Appl
Sequence 66, Appl
Sequence 438, App
Sequence 241, App
Sequence 14, Appl
Sequence 544, App
Sequence 12, Appl
Sequence 68, Appl
Sequence 68, Appl
Sequence 291, App
Sequence 24, Appl
Sequence 69, Appl
Sequence 5, Appl
Sequence 10, Appl
Sequence 102, App
Sequence 74, Appl
Sequence 103, App
Sequence 74, Appl
Sequence 103, App
Sequence 2, Appl
Sequence 3, Appl
Sequence 20, Appl
Sequence 18, Appl
Sequence 20, Appl
Sequence 18, Appl
Sequence 20, Appl
Sequence 68, Appl
Sequence 85, Appl
Sequence 20, Appl
Sequence 18, Appl
Sequence 16, Appl
Sequence 22, Appl
Sequence 16, Appl
Sequence 1, Appl
Sequence 16, Appl
Sequence 1, Appl
Sequence 16, Appl
Sequence 26, Appl
Sequence 22, Appl
Sequence 3595, Ap
Sequence 22, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 57, Appl
Sequence 31, Appl
Sequence 31, Appl
Sequence 31, Appl

524	6	0.5	9	1	US-08-076-092-62	Sequence 62, Appl	597	6	0.5	15	2	US-08-475-352-9	Sequence 9, Appl1
525	6	0.5	9	1	US-08-076-092-65	Sequence 65, Appl	598	6	0.5	15	4	US-08-602-999A-326	Sequence 326, App
526	6	0.5	9	1	US-08-421-702A-14	Sequence 14, Appl	599	6	0.5	15	4	US-08-602-999A-327	Sequence 327, App
527	6	0.5	9	1	US-08-303-052A-14	Sequence 14, Appl	600	6	0.5	15	4	US-08-602-999A-328	Sequence 328, App
528	6	0.5	9	1	US-08-421-698A-14	Sequence 14, Appl	601	6	0.5	15	4	US-08-602-999A-329	Sequence 329, App
529	6	0.5	9	1	US-08-421-697A-14	Sequence 14, Appl	602	6	0.5	15	4	US-08-602-999A-351	Sequence 351, App
530	6	0.5	9	1	US-08-421-698A-14	Sequence 14, Appl	603	6	0.5	15	4	US-08-468-337-19	Sequence 19, Appl
531	6	0.5	9	2	US-08-421-695A-14	Sequence 14, Appl	604	6	0.5	15	4	US-09-019-346A-12	Sequence 12, Appl
532	6	0.5	9	2	US-08-340-283-72	Sequence 72, Appl	605	6	0.5	15	4	US-09-500-124-326	Sequence 326, App
533	6	0.5	9	2	US-08-340-283-85	Sequence 85, Appl	606	6	0.5	15	4	US-09-500-124-327	Sequence 327, App
534	6	0.5	9	2	US-08-730-486-62	Sequence 62, Appl	607	6	0.5	15	4	US-09-500-124-328	Sequence 328, App
535	6	0.5	9	2	US-08-730-486-65	Sequence 65, Appl	608	6	0.5	15	4	US-09-500-124-329	Sequence 329, App
536	6	0.5	9	2	US-08-990-379-18	Sequence 18, Appl	609	6	0.5	15	4	US-09-500-124-351	Sequence 351, App
537	6	0.5	9	5	PCT-US95-04741-14	Sequence 14, Appl	610	6	0.5	15	5	PCT-US93-06751-37	Sequence 37, Appl
538	6	0.5	10	1	US-08-076-092-60	Sequence 60, Appl	611	6	0.5	15	5	PCT-US94-05796-9	Sequence 9, Appl1
539	6	0.5	10	1	US-08-076-092-63	Sequence 63, Appl	612	6	0.5	16	1	US-08-218-025A-12	Sequence 12, Appl
540	6	0.5	10	1	US-08-440-504A-10	Sequence 10, Appl	613	6	0.5	16	1	US-07-942-245-43	Sequence 43, Appl
541	6	0.5	10	2	US-08-716-317-32	Sequence 32, Appl	614	6	0.5	16	1	US-07-942-245-44	Sequence 44, Appl
542	6	0.5	10	2	US-08-747-137-80	Sequence 80, Appl	615	6	0.5	16	1	US-07-942-245-48	Sequence 48, Appl
543	6	0.5	10	2	US-08-730-486-60	Sequence 60, Appl	616	6	0.5	16	1	US-07-942-245-50	Sequence 50, Appl
544	6	0.5	10	2	US-08-730-486-63	Sequence 63, Appl	617	6	0.5	16	2	US-08-637-759B-465	Sequence 465, App
545	6	0.5	10	3	US-08-526-136-18	Sequence 18, Appl	618	6	0.5	16	3	US-08-871-355A-465	Sequence 465, App
546	6	0.5	10	4	US-09-319-730-8	Sequence 8, Appl1	619	6	0.5	16	3	US-09-037-190-27	Sequence 27, Appl
547	6	0.5	10	4	US-09-319-730-9	Sequence 9, Appl1	620	6	0.5	16	3	US-09-037-192-27	Sequence 27, Appl
548	6	0.5	10	5	PCT-US95-03236-17	Sequence 17, Appl	621	6	0.5	16	3	US-09-037-143-27	Sequence 27, Appl
549	6	0.5	11	1	US-08-076-092-5	Sequence 5, Appl	622	6	0.5	16	4	US-09-049-691-27	Sequence 27, Appl
550	6	0.5	11	1	US-08-076-092-59	Sequence 59, Appl	623	6	0.5	16	4	US-08-260-174-27	Sequence 27, Appl
551	6	0.5	11	1	US-08-076-092-61	Sequence 61, Appl	624	6	0.5	16	4	US-09-338-128A-27	Sequence 27, Appl
552	6	0.5	11	2	US-08-705-660-1	Sequence 1, Appl1	c 625	6	0.5	16	4	US-09-201-945-465	Sequence 465, App
553	6	0.5	11	2	US-08-730-486-5	Sequence 5, Appl1	626	6	0.5	16	4	US-09-232-346-27	Sequence 27, Appl
554	6	0.5	11	2	US-08-730-486-59	Sequence 59, Appl	627	6	0.5	16	4	US-09-037-192-27	Sequence 27, Appl
555	6	0.5	11	2	US-08-730-486-61	Sequence 61, Appl	628	6	0.5	16	4	US-09-288-339-24	Sequence 24, Appl
556	6	0.5	11	3	US-08-989-045-1	Sequence 1, Appl1	629	6	0.5	16	6	5378805-3	Patent No. 5378805
557	6	0.5	11	4	US-08-817-177-4	Sequence 4, Appl1	630	6	0.5	17	1	US-08-370-567-9	Sequence 9, Appl1
558	6	0.5	11	4	US-09-227-357-596	Sequence 596, Appl	631	6	0.5	17	1	US-08-438-759-9	Sequence 9, Appl1
559	6	0.5	11	4	US-09-288-339-14	Sequence 14, Appl	632	6	0.5	17	1	US-08-538-911-21	Sequence 21, Appl
560	6	0.5	11	5	PCT-US95-12686-4	Sequence 4, Appl1	633	6	0.5	17	4	US-09-308-022-3	Sequence 3, Appl1
561	6	0.5	12	1	US-07-939-501A-5	Sequence 5, Appl1	634	6	0.5	17	5	US-09-025-769B-229	Sequence 229, App
562	6	0.5	12	1	US-08-076-092-58	Sequence 58, Appl	635	6	0.5	17	5	PCT-US94-05591-21	Sequence 21, Appl
563	6	0.5	12	1	US-08-240-514-24	Sequence 24, Appl	636	6	0.5	17	5	PCT-US94-05684-9	Sequence 9, Appl1
564	6	0.5	12	1	US-08-548-540-129	Sequence 129, App	637	6	0.5	18	1	US-08-488-212A-30	Sequence 30, Appl
565	6	0.5	12	2	US-08-612-302A-24	Sequence 24, Appl	638	6	0.5	18	1	US-08-207-169A-7	Sequence 7, Appl1
566	6	0.5	12	2	US-08-406-330-24	Sequence 24, Appl	639	6	0.5	18	1	US-08-338-127-2	Sequence 2, Appl1
567	6	0.5	12	2	US-08-392-973A-2	Sequence 2, Appl1	640	6	0.5	18	2	US-08-320-306-30	Sequence 30, Appl
568	6	0.5	12	2	US-08-556-597-24	Sequence 24, Appl	641	6	0.5	18	2	US-08-488-209B-30	Sequence 30, Appl
569	6	0.5	12	2	US-08-730-486-58	Sequence 58, Appl	642	6	0.5	18	2	US-08-408-011-30	Sequence 30, Appl
570	6	0.5	12	4	US-09-258-754-89	Sequence 89, Appl	c 643	6	0.5	18	2	US-09-017-205-47	Sequence 47, Appl
571	6	0.5	12	4	US-08-448-398-9	Sequence 9, Appl1	644	6	0.5	18	2	US-09-017-205-48	Sequence 48, Appl
572	6	0.5	12	4	US-09-042-107-89	Sequence 89, Appl	645	6	0.5	19	1	US-07-596-081A-28	Sequence 28, Appl
573	6	0.5	12	5	PCT-US93-12679-26	Sequence 26, Appl	646	6	0.5	19	2	US-08-019-073-30	Sequence 30, Appl
574	6	0.5	12	5	PCT-US96-09809-129	Sequence 129, App	647	6	0.5	19	2	US-08-764-640-158	Sequence 158, App
575	6	0.5	13	3	US-09-037-190-23	Sequence 23, Appl	648	6	0.5	19	2	US-08-483-636-8	Sequence 8, Appl1
576	6	0.5	13	3	US-09-037-192-23	Sequence 23, Appl	649	6	0.5	19	2	US-08-483-632-8	Sequence 8, Appl1
577	6	0.5	13	4	US-09-037-143-23	Sequence 23, Appl	650	6	0.5	19	3	US-09-100-414B-17	Sequence 17, Appl
578	6	0.5	13	4	US-09-049-691-23	Sequence 23, Appl	651	6	0.5	19	3	US-08-973-225-158	Sequence 158, App
579	6	0.5	13	4	US-08-602-999A-83	Sequence 83, Appl	c 652	6	0.5	19	3	US-09-244-298A-158	Sequence 158, App
580	6	0.5	13	4	US-08-260-174-23	Sequence 23, Appl	653	6	0.5	19	4	US-09-303-323-17	Sequence 17, Appl
581	6	0.5	13	4	US-08-278-865-83	Sequence 83, Appl	654	6	0.5	19	4	US-08-928-213B-132	Sequence 132, App
582	6	0.5	13	4	US-09-319-730-2	Sequence 2, Appl1	c 655	6	0.5	19	4	US-09-516-704-158	Sequence 158, App
583	6	0.5	13	4	US-09-338-128A-23	Sequence 23, Appl	656	6	0.5	19	4	US-09-441-502B-11	Sequence 11, Appl
584	6	0.5	13	4	US-09-232-346-23	Sequence 23, Appl	657	6	0.5	19	4	US-09-549-090-158	Sequence 158, App
585	6	0.5	13	4	US-09-037-192-23	Sequence 23, Appl	658	6	0.5	19	5	PCT-US94-01768-30	Sequence 30, Appl
586	6	0.5	13	4	US-08-182-967-15	Sequence 15, Appl	659	6	0.5	20	1	US-07-700-526-15	Sequence 15, Appl
587	6	0.5	13	4	US-09-500-124-83	Sequence 83, Appl	c 660	6	0.5	20	1	US-07-678-974D-43	Sequence 43, Appl
588	6	0.5	14	1	US-08-320-373-9	Sequence 9, Appl1	661	6	0.5	20	1	US-07-678-974D-53	Sequence 53, Appl
589	6	0.5	14	2	US-08-442-461D-4	Sequence 4, Appl1	662	6	0.5	20	1	US-08-338-127-1	Sequence 1, Appl1
590	6	0.5	14	4	US-09-426-680-4	Sequence 4, Appl1	663	6	0.5	20	1	US-08-222-851-40	Sequence 40, Appl
591	6	0.5	14	4	US-09-608-285A-44	Sequence 44, Appl	664	6	0.5	20	1	US-08-732-751-10	Sequence 10, Appl
592	6	0.5	15	1	US-07-978-895-9	Sequence 9, Appl1	665	6	0.5	20	2	US-08-934-915-10	Sequence 10, Appl
593	6	0.5	15	1	US-08-067-387-9	Sequence 9, Appl1	666	6	0.5	20	2	US-08-934-915-17	Sequence 17, Appl
594	6	0.5	15	1	US-08-320-373-50	Sequence 50, Appl	667	6	0.5	20	2	US-08-934-915-40	Sequence 40, Appl
595	6	0.5	15	1	US-08-237-418-19	Sequence 19, Appl	668	6	0.5	20	2	US-08-934-915-141	Sequence 141, App
596	6	0.5	15	2	US-08-473-119-9	Sequence 9, Appl1	669	6	0.5	20	2	US-08-934-915-170	Sequence 170, App

c 670	6	0.5	20	2	US-08-945-168-49	Sequence 49, Appl	c 743	6	0.5	27	3	US-08-834-306-46	Sequence 46, Appl
c 671	6	0.5	20	2	US-08-945-168-59	Sequence 59, Appl	c 744	6	0.5	27	4	US-08-993-674A-46	Sequence 46, Appl
c 672	6	0.5	20	2	US-09-010-928B-7	Sequence 7, Appl	c 745	6	0.5	27	4	US-09-406-781-14	Sequence 14, Appl
c 673	6	0.5	20	3	US-08-460-576-15	Sequence 15, Appl	c 746	6	0.5	27	4	US-09-256-976-46	Sequence 46, Appl
c 674	6	0.5	20	4	US-08-602-999A-110	Sequence 110, App	c 747	6	0.5	28	1	US-08-485-692-13	Sequence 13, Appl
c 675	6	0.5	20	4	US-08-602-999A-116	Sequence 116, App	c 748	6	0.5	28	1	US-08-732-751-11	Sequence 11, Appl
c 676	6	0.5	20	4	US-08-928-213B-119	Sequence 119, App	c 749	6	0.5	28	1	US-08-419-519-13	Sequence 13, Appl
c 677	6	0.5	20	4	US-08-928-213B-123	Sequence 123, App	c 750	6	0.5	28	4	US-09-177-249-97	Sequence 97, Appl
c 678	6	0.5	20	4	US-08-928-213B-124	Sequence 124, App	c 751	6	0.5	28	4	US-08-918-288-1	Sequence 1, Appl
c 679	6	0.5	20	4	US-08-928-213B-125	Sequence 125, App	c 752	6	0.5	28	4	US-09-282-357-1	Sequence 1, Appl
c 680	6	0.5	20	4	US-08-928-213B-127	Sequence 127, App	c 753	6	0.5	28	4	US-09-604-871-3	Sequence 3, Appl
c 681	6	0.5	20	4	US-08-928-213B-128	Sequence 128, App	c 754	6	0.5	29	1	US-07-960-510-4	Sequence 4, Appl
c 682	6	0.5	20	4	US-08-928-213B-129	Sequence 129, App	c 755	6	0.5	30	1	US-08-732-751-7	Sequence 7, Appl
c 683	6	0.5	20	4	US-08-928-213B-130	Sequence 130, App	c 756	6	0.5	30	2	US-08-753-829A-14	Sequence 14, Appl
c 684	6	0.5	20	4	US-08-817-895-12	Sequence 12, Appl	c 757	6	0.5	30	2	US-08-716-317-16	Sequence 16, Appl
c 685	6	0.5	20	4	US-09-500-124-110	Sequence 110, App	c 758	6	0.5	30	2	US-08-934-915-42	Sequence 42, Appl
c 686	6	0.5	20	4	US-09-500-124-116	Sequence 116, App	c 759	6	0.5	30	2	US-08-934-915-190	Sequence 190, App
c 687	6	0.5	20	4	US-08-651-650-19	Sequence 19, Appl	c 760	6	0.5	30	2	US-08-934-915-193	Sequence 193, App
c 688	6	0.5	20	5	PCT-US92-03132-15	Sequence 15, Appl	c 761	6	0.5	30	2	US-08-031-538-50	Sequence 50, Appl
c 689	6	0.5	20	5	PCT-US94-12985-2	Sequence 2, Appl	c 762	6	0.5	30	4	US-09-037-179B-14	Sequence 14, Appl
c 690	6	0.5	20	5	PCT-US94-12985-8	Sequence 8, Appl	c 763	6	0.5	30	4	US-09-037-179B-17	Sequence 17, Appl
c 691	6	0.5	21	1	US-07-969-931-2	Sequence 2, Appl	c 764	6	0.5	31	1	US-07-596-081A-1	Sequence 1, Appl
c 692	6	0.5	21	1	US-07-855-417A-2	Sequence 2, Appl	c 765	6	0.5	31	1	US-07-596-081A-24	Sequence 24, Appl
c 693	6	0.5	21	2	US-08-894-339-8	Sequence 8, Appl	c 766	6	0.5	31	1	US-08-732-751-3	Sequence 3, Appl
c 694	6	0.5	21	2	US-08-454-236-9	Sequence 9, Appl	c 767	6	0.5	31	1	US-08-732-751-4	Sequence 4, Appl
c 695	6	0.5	21	2	US-09-092-770-16	Sequence 16, Appl	c 768	6	0.5	31	2	US-08-455-625-6	Sequence 6, Appl
c 696	6	0.5	21	4	US-09-222-851-16	Sequence 16, Appl	c 769	6	0.5	31	2	US-08-455-625-33	Sequence 33, Appl
c 697	6	0.5	21	4	US-09-306-044-8	Sequence 8, Appl	c 770	6	0.5	31	3	US-09-100-414B-55	Sequence 55, Appl
c 698	6	0.5	21	4	US-09-227-357-284	Sequence 284, App	c 771	6	0.5	31	4	US-08-455-685-6	Sequence 6, Appl
c 699	6	0.5	22	1	US-07-596-081A-7	Sequence 7, Appl	c 772	6	0.5	32	1	US-08-455-685-33	Sequence 33, Appl
c 700	6	0.5	22	1	US-08-103-445-15	Sequence 15, Appl	c 773	6	0.5	32	1	US-09-303-323-55	Sequence 55, Appl
c 701	6	0.5	22	1	US-08-461-690B-15	Sequence 15, Appl	c 774	6	0.5	32	4	US-08-060-988A-6	Sequence 6, Appl
c 702	6	0.5	22	1	US-08-241-853-37	Sequence 37, Appl	c 775	6	0.5	32	4	US-08-060-988A-33	Sequence 33, Appl
c 703	6	0.5	22	2	US-08-850-917-37	Sequence 37, Appl	c 776	6	0.5	32	5	PCT-US94-05142-6	Sequence 6, Appl
c 704	6	0.5	22	2	US-08-525-742-23	Sequence 23, Appl	c 777	6	0.5	32	5	PCT-US94-05142-33	Sequence 33, Appl
c 705	6	0.5	22	2	US-08-194-981B-18	Sequence 18, Appl	c 778	6	0.5	32	1	US-08-152-488-9	Sequence 9, Appl
c 706	6	0.5	22	3	US-08-783-853A-39	Sequence 39, Appl	c 779	6	0.5	32	1	US-08-303-025-9	Sequence 9, Appl
c 707	6	0.5	22	4	US-09-406-781-41	Sequence 41, Appl	c 780	6	0.5	32	1	US-08-677-304-9	Sequence 9, Appl
c 708	6	0.5	22	4	US-09-344-050-39	Sequence 39, Appl	c 781	6	0.5	32	2	US-08-716-317-21	Sequence 21, Appl
c 709	6	0.5	22	4	US-09-149-476-320	Sequence 320, App	c 782	6	0.5	32	2	US-08-436-703B-14	Sequence 14, Appl
c 710	6	0.5	22	4	US-09-149-476-506	Sequence 506, App	c 783	6	0.5	32	4	US-09-252-586-15	Sequence 15, Appl
c 711	6	0.5	22	4	US-09-149-476-507	Sequence 507, App	c 784	6	0.5	33	1	US-07-596-081A-6	Sequence 6, Appl
c 712	6	0.5	22	6	5204326-76	Patent No. 5204326	c 785	6	0.5	33	2	US-08-146-028-32	Sequence 32, Appl
c 713	6	0.5	23	2	US-08-460-890A-56	Sequence 56, Appl	c 786	6	0.5	33	2	US-08-146-028-171	Sequence 171, App
c 714	6	0.5	23	3	US-08-783-853A-24	Sequence 24, Appl	c 787	6	0.5	33	4	US-08-723-425A-32	Sequence 32, Appl
c 715	6	0.5	23	3	US-08-783-853A-76	Sequence 76, Appl	c 788	6	0.5	33	4	US-08-723-425A-171	Sequence 171, App
c 716	6	0.5	23	3	US-08-167-641C-56	Sequence 56, Appl	c 789	6	0.5	33	4	US-09-112-206-32	Sequence 32, Appl
c 717	6	0.5	23	4	US-08-460-971A-56	Sequence 56, Appl	c 790	6	0.5	33	4	US-09-112-206-171	Sequence 171, App
c 718	6	0.5	23	4	US-08-462-040-56	Sequence 56, Appl	c 791	6	0.5	33	4	US-08-857-046A-31	Sequence 31, Appl
c 719	6	0.5	23	4	US-09-344-050-24	Sequence 24, Appl	c 792	6	0.5	33	4	US-09-149-476-447	Sequence 447, App
c 720	6	0.5	23	4	US-09-344-050-76	Sequence 76, Appl	c 793	6	0.5	33	4	US-09-149-476-447	Sequence 447, App
c 721	6	0.5	23	4	US-08-835-281-5	Sequence 5, Appl	c 794	6	0.5	34	4	US-09-570-921-126	Sequence 126, App
c 722	6	0.5	23	4	US-08-835-281-6	Sequence 6, Appl	c 795	6	0.5	34	1	US-07-939-501A-21	Sequence 21, Appl
c 723	6	0.5	23	6	5451527-3	Patent No. 5451527	c 796	6	0.5	34	1	US-08-700-749A-3	Sequence 3, Appl
c 724	6	0.5	24	1	US-09-217-751-5	Sequence 5, Appl	c 797	6	0.5	34	2	US-08-716-317-26	Sequence 26, Appl
c 725	6	0.5	24	4	US-09-937-306B-24	Sequence 24, Appl	c 798	6	0.5	34	2	US-08-454-236-6	Sequence 6, Appl
c 726	6	0.5	25	2	US-08-942-423-26	Sequence 26, Appl	c 799	6	0.5	34	3	US-09-020-684-3	Sequence 3, Appl
c 727	6	0.5	25	4	US-09-227-357-320	Sequence 320, App	c 800	6	0.5	34	3	US-09-020-684-3	Sequence 3, Appl
c 728	6	0.5	25	4	US-08-826-134-16	Sequence 16, Appl	c 801	6	0.5	34	3	US-09-020-685-3	Sequence 3, Appl
c 729	6	0.5	25	4	US-08-826-134-27	Sequence 27, Appl	c 802	6	0.5	34	4	US-09-020-683-3	Sequence 3, Appl
c 730	6	0.5	25	5	PCT-US93-05325-10	Sequence 10, Appl	c 803	6	0.5	34	4	US-09-227-357-321	Sequence 321, App
c 731	6	0.5	26	1	US-07-942-245-378	Sequence 378, App	c 804	6	0.5	34	6	5451527-4	Patent No. 5451527
c 732	6	0.5	26	2	US-08-620-151-83	Sequence 83, Appl	c 805	6	0.5	35	2	US-08-867-087B-60	Sequence 60, Appl
c 733	6	0.5	26	3	US-08-755-587-217	Sequence 217, App	c 806	6	0.5	35	2	US-08-867-087B-61	Sequence 61, Appl
c 734	6	0.5	26	4	US-08-602-999A-51	Sequence 51, Appl	c 807	6	0.5	35	3	US-09-053-197A-63	Sequence 63, Appl
c 735	6	0.5	26	4	US-09-500-124-51	Sequence 51, Appl	c 808	6	0.5	35	4	US-08-905-223-331	Sequence 331, App
c 736	6	0.5	27	1	US-08-786-748A-83	Sequence 83, Appl	c 809	6	0.5	35	4	US-09-085-761A-68	Sequence 68, Appl
c 737	6	0.5	27	1	US-08-786-748A-84	Sequence 84, Appl	c 810	6	0.5	35	4	US-09-314-268-117	Sequence 117, App
c 738	6	0.5	27	1	US-08-732-751-12	Sequence 12, Appl	c 811	6	0.5	36	1	US-08-237-418-20	Sequence 20, Appl
c 739	6	0.5	27	1	US-08-557-309B-46	Sequence 46, Appl	c 812	6	0.5	36	4	US-08-468-337-20	Sequence 20, Appl
c 740	6	0.5	27	2	US-08-932-682-83	Sequence 83, Appl	c 813	6	0.5	36	4	US-09-215-212-8	Sequence 8, Appl
c 741	6	0.5	27	2	US-08-932-682-84	Sequence 84, Appl	c 814	6	0.5	36	4	US-09-092-315-19	Sequence 19, Appl
c 742	6	0.5	27	3	US-08-654-623-23	Sequence 23, Appl	c 815	6	0.5	36	4	US-08-469-260A-414	Sequence 414, App

c 816	6	0.5	36	6	5451527-5	Patent No. 5451527	889	6	0.5	45	4	US-09-314-268-102	Sequence 102, App
c 817	6	0.5	36	6	5451527-6	Patent No. 5451527	c 890	6	0.5	45	4	US-09-314-268-102	Sequence 102, App
c 818	6	0.5	37	2	US-08-709-924-25	Sequence 25, Appl	c 891	6	0.5	45	4	US-09-314-268-103	Sequence 103, App
c 819	6	0.5	37	2	US-08-709-925-25	Sequence 25, Appl	c 892	6	0.5	45	6	5451527-1	Patent No. 5451527
c 820	6	0.5	37	4	US-08-709-948-25	Sequence 25, Appl	c 893	6	0.5	46	4	US-09-004-406C-14	Sequence 14, Appl
c 821	6	0.5	37	4	US-09-149-476-694	Sequence 694, App	c 894	6	0.5	46	4	US-08-469-260A-256	Sequence 256, App
c 822	6	0.5	38	1	US-08-176-500-25	Sequence 25, Appl	c 895	6	0.5	47	4	US-09-227-357-637	Sequence 637, App
c 823	6	0.5	38	1	US-08-176-500-131	Sequence 131, App	c 896	6	0.5	48	2	US-08-637-759B-199	Sequence 199, App
c 824	6	0.5	38	1	US-08-471-052A-25	Sequence 25, Appl	c 897	6	0.5	48	3	US-08-871-355A-199	Sequence 199, App
c 825	6	0.5	38	1	US-08-471-052A-131	Sequence 131, App	c 898	6	0.5	48	4	US-09-201-945-199	Sequence 199, App
c 826	6	0.5	38	1	US-08-189-331-25	Sequence 25, Appl	c 899	6	0.5	49	2	US-08-765-179B-7	Sequence 7, Appl
c 827	6	0.5	38	1	US-08-189-331-131	Sequence 131, App	c 900	6	0.5	49	4	US-09-227-357-595	Sequence 595, App
c 828	6	0.5	38	2	US-08-471-939-25	Sequence 25, Appl	c 901	6	0.5	49	4	US-09-065-383-31	Sequence 31, Appl
c 829	6	0.5	38	2	US-08-471-939-131	Sequence 131, App	c 902	6	0.5	50	4	US-09-314-268-159	Sequence 159, App
c 830	6	0.5	38	2	US-08-471-939-131	Sequence 25, Appl	c 903	6	0.5	51	1	US-08-445-909A-18	Sequence 18, Appl
c 831	6	0.5	38	2	US-08-471-800-131	Sequence 131, App	c 904	6	0.5	51	4	US-09-230-637-46	Sequence 46, Appl
c 832	6	0.5	38	2	US-08-488-161-14	Sequence 14, Appl	c 905	6	0.5	52	1	US-08-346-849-8	Sequence 8, Appl
c 833	6	0.5	38	2	US-08-471-068-25	Sequence 25, Appl	c 906	6	0.5	52	2	US-08-293-284A-8	Sequence 8, Appl
c 834	6	0.5	38	2	US-08-471-068-131	Sequence 131, App	c 907	6	0.5	52	3	US-08-917-299-32	Sequence 32, Appl
c 835	6	0.5	38	2	US-08-726-306A-88	Sequence 88, Appl	c 908	6	0.5	52	4	US-09-422-662-32	Sequence 32, Appl
c 836	6	0.5	38	3	US-09-273-685-14	Sequence 14, Appl	c 909	6	0.5	52	4	US-09-153-599A-4	Sequence 4, Appl
c 837	6	0.5	38	5	PCR-US95-1193A-14	Sequence 14, Appl	c 910	6	0.5	53	2	US-08-852-401-5	Sequence 5, Appl
c 838	6	0.5	39	2	US-08-657-392-29	Sequence 29, Appl	c 911	6	0.5	53	4	US-09-149-476-448	Sequence 448, App
c 839	6	0.5	39	2	US-08-488-161-33	Sequence 33, Appl	c 912	6	0.5	54	4	US-08-905-223-427	Sequence 427, App
c 840	6	0.5	39	3	US-09-273-685-33	Sequence 33, Appl	c 913	6	0.5	54	4	US-09-306-446C-13	Sequence 13, Appl
c 841	6	0.5	39	4	US-09-314-268-155	Sequence 155, App	c 914	6	0.5	55	3	US-08-814-836-7	Sequence 7, Appl
c 842	6	0.5	39	4	US-09-314-268-156	Sequence 156, App	c 915	6	0.5	55	4	US-09-314-268-147	Sequence 147, App
c 843	6	0.5	39	4	US-09-384-302A-7	Sequence 7, Appl	c 916	6	0.5	56	4	US-09-227-357-577	Sequence 577, App
c 844	6	0.5	39	5	PCR-US94-02539-29	Sequence 29, Appl	c 917	6	0.5	57	1	US-08-137-614A-17	Sequence 17, Appl
c 845	6	0.5	39	5	PCR-US95-11934-33	Sequence 33, Appl	c 918	6	0.5	57	3	US-08-814-836-17	Sequence 17, Appl
c 846	6	0.5	39	6	5451527-8	Patent No. 5451527	c 919	6	0.5	58	1	US-08-406-809-3	Sequence 3, Appl
c 847	6	0.5	39	6	5451527-12	Patent No. 5451527	c 920	6	0.5	58	1	US-08-406-809-4	Sequence 4, Appl
c 848	6	0.5	40	1	US-08-188-228-18	Sequence 18, Appl	c 921	6	0.5	58	1	US-08-401-908-31	Sequence 31, Appl
c 849	6	0.5	40	1	US-08-332-643-18	Sequence 18, Appl	c 922	6	0.5	58	4	US-09-461-697-173	Sequence 173, App
c 850	6	0.5	40	1	US-08-332-638-18	Sequence 18, Appl	c 923	6	0.5	58	4	US-09-319-730-17	Sequence 17, Appl
c 851	6	0.5	40	2	US-08-194-981B-14	Sequence 14, Appl	c 924	6	0.5	59	1	US-08-287-959-21	Sequence 21, Appl
c 852	6	0.5	40	2	US-08-194-981B-15	Sequence 15, Appl	c 925	6	0.5	59	2	US-08-716-317-2	Sequence 2, Appl
c 853	6	0.5	40	3	US-09-037-524-4	Sequence 4, Appl	c 926	6	0.5	59	3	US-08-657-983A-4	Sequence 4, Appl
c 854	6	0.5	40	4	US-08-988-856B-7	Sequence 7, Appl	c 927	6	0.5	59	4	US-09-036-549A-4	Sequence 4, Appl
c 855	6	0.5	40	4	US-08-908-371B-10	Sequence 10, Appl	c 928	6	0.5	59	4	US-08-924-629C-30	Sequence 30, Appl
c 856	6	0.5	41	1	US-07-709-091-3	Sequence 3, Appl	c 929	6	0.5	59	4	US-09-082-358B-50	Sequence 50, Appl
c 857	6	0.5	41	1	US-07-715-752A-3	Sequence 3, Appl	c 930	6	0.5	60	2	US-08-117-952-788	Sequence 788, App
c 858	6	0.5	41	1	US-08-104-862-3	Sequence 3, Appl	c 931	6	0.5	60	2	US-08-607-412-2	Sequence 2, Appl
c 859	6	0.5	41	1	US-08-162-178-3	Sequence 3, Appl	c 932	6	0.5	60	4	US-09-227-357-307	Sequence 307, App
c 860	6	0.5	41	1	US-08-122-520C-2	Sequence 2, Appl	c 933	6	0.5	60	4	US-09-227-357-317	Sequence 317, App
c 861	6	0.5	41	1	US-08-865-773-4	Sequence 4, Appl	c 934	6	0.5	60	4	US-09-227-357-518	Sequence 518, App
c 862	6	0.5	41	2	US-08-194-981B-17	Sequence 17, Appl	c 935	6	0.5	60	4	US-08-924-629C-31	Sequence 31, Appl
c 863	6	0.5	41	4	US-09-174-465D-8	Sequence 8, Appl	c 936	6	0.5	60	4	US-09-336-536-36	Sequence 36, Appl
c 864	6	0.5	41	4	US-09-424-127-4	Sequence 4, Appl	c 937	6	0.5	61	4	US-09-319-730-15	Sequence 15, Appl
c 865	6	0.5	41	4	US-09-314-268-104	Sequence 104, App	c 938	6	0.5	61	4	US-08-858-207A-538	Sequence 538, App
c 866	6	0.5	41	4	US-08-936-165A-290	Sequence 290, App	c 939	6	0.5	63	6	5164482-9	Patent No. 5164482
c 867	6	0.5	41	4	US-09-599-564A-8	Sequence 8, Appl	c 940	6	0.5	64	1	US-08-470-202-39	Sequence 39, Appl
c 868	6	0.5	41	4	US-09-201-227A-27	Sequence 27, Appl	c 941	6	0.5	64	1	US-08-471-770-39	Sequence 39, Appl
c 869	6	0.5	41	5	PCR-US92-05101-3	Sequence 3, Appl	c 942	6	0.5	64	2	US-08-292-870-1	Sequence 1, Appl
c 870	6	0.5	41	6	5451527-9	Patent No. 5451527	c 943	6	0.5	64	2	US-08-468-059-39	Sequence 39, Appl
c 871	6	0.5	41	6	5451527-10	Patent No. 5451527	c 944	6	0.5	64	4	US-09-109-916-39	Sequence 39, Appl
c 872	6	0.5	41	6	5451527-11	Patent No. 5451527	c 945	6	0.5	64	4	US-09-134-001C-4846	Sequence 4846, Ap
c 873	6	0.5	41	6	5451527-13	Patent No. 5451527	c 946	6	0.5	64	4	US-08-469-260A-384	Sequence 384, App
c 874	6	0.5	41	6	5451527-14	Patent No. 5451527	c 947	6	0.5	66	4	US-09-227-357-160	Sequence 160, App
c 875	6	0.5	42	1	US-07-651-710A-23	Sequence 23, App	c 948	6	0.5	67	2	US-08-273-146-67	Sequence 67, Appl
c 876	6	0.5	42	3	US-08-924-330A-3	Sequence 3, Appl	c 949	6	0.5	67	4	US-09-331-254-2	Sequence 2, Appl
c 877	6	0.5	42	4	US-09-138-721-3	Sequence 3, Appl	c 950	6	0.5	68	1	US-07-756-250-10	Sequence 10, Appl
c 878	6	0.5	42	4	US-09-156-580-9	Sequence 9, Appl	c 951	6	0.5	68	2	US-08-511-485-16	Sequence 16, Appl
c 879	6	0.5	43	4	US-08-905-223-293	Sequence 293, App	c 952	6	0.5	68	4	US-09-220-528-50	Sequence 50, Appl
c 880	6	0.5	43	4	US-09-156-579C-11	Sequence 11, Appl	c 953	6	0.5	69	2	US-08-997-080-184	Sequence 184, App
c 881	6	0.5	43	4	US-08-908-371B-12	Sequence 12, Appl	c 954	6	0.5	69	2	US-08-997-362-184	Sequence 184, App
c 882	6	0.5	44	2	US-08-484-434C-24	Sequence 24, Appl	c 955	6	0.5	69	4	US-09-095-855-184	Sequence 184, App
c 883	6	0.5	44	3	US-08-707-399B-18	Sequence 18, Appl	c 956	6	0.5	69	4	US-09-324-542-184	Sequence 184, App
c 884	6	0.5	45	2	US-08-530-290-21	Sequence 21, Appl	c 957	6	0.5	69	4	US-09-134-001C-3959	Sequence 3959, Ap
c 885	6	0.5	45	3	US-09-100-600A-56	Sequence 56, Appl	c 958	6	0.5	69	4	US-09-205-426-184	Sequence 184, App
c 886	6	0.5	45	3	US-08-236-886-2	Sequence 2, Appl	c 959	6	0.5	70	4	US-09-367-953B-68	Sequence 68, Appl
c 887	6	0.5	45	4	US-08-857-076-62	Sequence 62, Appl	c 960	6	0.5	71	2	US-08-972-008-4	Sequence 4, Appl
c 888	6	0.5	45	4	US-08-857-076-92	Sequence 92, Appl	c 961	6	0.5	71	4	US-09-134-001C-4418	Sequence 4418, Ap

```
c 962      6      0.5      71      4      US-08-267-409-4      Sequence 4, Appli
c 963      6      0.5      72      1      US-08-439-818A-21      Sequence 21, Appl
c 964      6      0.5      72      1      US-08-542-363-17      Sequence 17, Appl
c 965      6      0.5      72      2      US-08-751-965-21      Sequence 21, Appl
c 966      6      0.5      72      2      US-08-738-975-21      Sequence 21, Appl
c 967      6      0.5      72      2      US-08-728-826-21      Sequence 21, Appl
c 968      6      0.5      72      2      US-08-530-290-19      Sequence 19, Appl
c 969      6      0.5      72      3      US-08-808-599A-21      Sequence 21, Appl
c 970      6      0.5      72      4      US-09-100-089-17      Sequence 17, Appl
c 971      6      0.5      72      4      US-08-905-223-314      Sequence 314, App
c 972      6      0.5      72      4      US-09-670-827-17      Sequence 17, Appl
c 973      6      0.5      73      4      US-09-509-902A-8       Sequence 8, Appli
c 974      6      0.5      73      4      US-08-469-260A-510     Sequence 510, App
c 975      6      0.5      74      2      US-08-343-443B-11      Sequence 11, Appl
c 976      6      0.5      74      4      US-09-337-227C-30      Sequence 30, Appl
c 977      6      0.5      75      1      US-07-881-075-12      Sequence 12, Appl
c 978      6      0.5      75      1      US-08-120-827-12      Sequence 12, Appl
c 979      6      0.5      75      1      US-08-478-675-12      Sequence 12, Appl
c 980      6      0.5      75      4      US-09-134-001C-3025    Sequence 3025, Ap
c 981      6      0.5      75      4      US-09-134-001C-5076    Sequence 5076, Ap
c 982      6      0.5      76      4      US-09-227-357-456      Sequence 456, App
c 983      6      0.5      76      4      US-09-281-766-12      Sequence 12, Appl
c 984      6      0.5      77      2      US-08-476-866-21      Sequence 21, Appl
c 985      6      0.5      78      2      US-08-469-412A-11      Sequence 11, Appl
c 986      6      0.5      78      4      US-09-021-715-11      Sequence 11, Appl
c 987      6      0.5      79      3      US-09-143-776-6        Sequence 6, Appli
c 988      6      0.5      80      2      US-08-332-562A-84      Sequence 84, Appl
c 989      6      0.5      81      4      US-09-249-542-2        Sequence 2, Appli
c 990      6      0.5      83      4      US-08-936-165A-410     Sequence 410, App
c 991      6      0.5      85      4      US-08-711-417C-190     Sequence 190, App
c 992      6      0.5      85      4      US-09-263-933-6        Sequence 6, Appli
c 993      6      0.5      85      4      US-09-263-933-13      Sequence 13, Appl
c 994      6      0.5      85      4      US-09-263-933-20      Sequence 20, Appl
c 995      6      0.5      86      4      US-08-858-207A-471     Sequence 471, App
c 996      6      0.5      87      4      US-08-711-417C-192     Sequence 192, App
c 997      6      0.5      87      4      US-08-711-417C-193     Sequence 193, App
c 998      6      0.5      87      4      US-09-220-528-115      Sequence 115, App
c 999      6      0.5      87      4      US-09-152-060-89      Sequence 89, Appl
c1000     6      0.5      88      1      US-08-469-202-20      Sequence 20, Appl
```

ALIGNMENTS

RESULT 1

```
us-09-149-476-477
; Sequence 477, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
```

```
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
```

/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,630
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,878
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,662
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,872
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,882
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,637
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,903
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,888
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,879
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,880
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,894
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,911
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,636
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,631
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,845
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,892
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/057,761
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/047,595
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,599
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,588
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,585
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,586
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,590
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,594
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,589
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,593
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,614
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,578
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,576
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/047,501
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,670
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/056,632
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,664
/ EARLIER FILING DATE: 1997-08-22

/ EARLIER APPLICATION NUMBER: 60/056,876
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,881
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,909
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,875
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,862
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,887
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,908
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/048,964
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/057,650
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/056,884
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/057,669
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/049,610
/ EARLIER FILING DATE: 1997-06-13
/ EARLIER APPLICATION NUMBER: 60/061,060
/ EARLIER FILING DATE: 1997-10-02

Alignment Scores:
Pred. No.: 75.4 Length: 192
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 4 Gaps: 0

US-09-818-990b-1 (1-3963) x US-09-149-476-477 (1-192)

QY 174 ATCTGGCTCGGCTCTTCAGC 151

Db 9 IleuAlaSerAlaSerPheSer 16

RESULT 2

US-09-500-569-2
; Sequence 2, Application us/09500569
; Patent No. 6329204
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: plant Caffeic acid 3-O-Methyltransferase Homologs
; FILE REFERENCE: BB1327 US NA
; CURRENT APPLICATION NUMBER: US/09/500,569
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,587
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (277)
US-09-500-569-2

Alignment Scores:
Pred. No.: 70.5 Length: 305
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-500-569-2 (1-305)

QY 2216 GGGCCACAGTTGCTGGGTGGTG 2193

Db 216 GYAlaThrValAlaAlaVal 223

RESULT 3

US-08-440-845D-5
; Sequence 5, Application US/08440845D
; Patent No. 5955329
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; APPLICANT: Kridl, J.
; APPLICANT: Dehesh, K.
; APPLICANT: Knauf, V.
; TITLE OF INVENTION: Engineering Plant Thioesterases For
; TITLE OF INVENTION: Altered Substrate Specificity.
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,845D
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 residues
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-440-845D-5

Alignment Scores:

Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-440-845D-5 (1-352)

QY 1382 CTCATCTCCTAAGTTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 4

US-08-868-458-5
; Sequence 5, Application US/08868458
; Patent No. 6150512
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; TITLE OF INVENTION: Engineering Plant Thioesterases And

US-08-868-458-5

Alignment Scores:

Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-868-458-5 (1-352)

QY 1382 CTCATCTCCTAAGTTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 5

US-09-134-262-2
; Sequence 2, Application US/09134262
; Patent No. 6365802
; GENERAL INFORMATION:
; APPLICANT: Calgene LLC
; TITLE OF INVENTION: Methods for Increasing Stearate Content in Soybean Oil
; FILE REFERENCE: Docket #17030
; CURRENT APPLICATION NUMBER: US/09/134,262
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: IBM PC; Windows NT 4.0; Microsoft Word For Windows 7.0a
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Garcinia mangostana
US-09-134-262-2

Alignment Scores:

Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-868-458-5 (1-352)

QY 1382 CTCATCTCCTAAGTTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 6

US-09-818-990B-1 (1-3963) x US-08-868-458-5 (1-352)

Alignment Scores:

Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 4 Gaps: 0

Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-134-262-2 (1-352)

QY 1382 CTCATCTCCTAGGTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 6

US-09-359-070-2
; Sequence 2, Application US/09359070
; Patent No. 6380462
; GENERAL INFORMATION:
; APPLICANT: Calgene LLC
; TITLE OF INVENTION: Methods for Increasing Stearate Content
; FILE REFERENCE: 17030/01/US
; CURRENT APPLICATION NUMBER: US/09/359,070
; CURRENT FILING DATE: 1999-07-22
; PRIOR APPLICATION NUMBER: 09/134,262
; PRIOR FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Garcinia mangostana
US-09-359-070-2

Alignment Scores:
Pred. No.: 69 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-359-070-2 (1-352)

QY 1382 CTCATCTCCTAGGTGAGTGGT 1405

Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 7

US-08-464-523B-32
; Sequence 32, Application US/08464523B
; Patent No. 5723761
; GENERAL INFORMATION:
; APPLICANT: Toni A. Voelker
; APPLICANT: Ling Yuan
; APPLICANT: Jean Kridl
; APPLICANT: Deborah Hawkins
; APPLICANT: Aubrey Jones
; TITLE OF INVENTION: Plant Acyl ACP Thioesterase
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1(a)

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,523B
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13131
; FILING DATE: 10-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/152,004
; FILING DATE: 10-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/261,695
; FILING DATE: 16-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 100-1WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-464-523B-32

Alignment Scores:
Pred. No.: 68.8 Length: 362
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-464-523B-32 (1-362)

QY 1382 CTCATCTCCTAGGTGAGTGGT 1405

Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 8

US-08-948-176-25
; Sequence 25, Application US/08948176
; Patent No. 5945585
; GENERAL INFORMATION:
; APPLICANT: HITZ, WILLIAM D.
; APPLICANT: YADAV, NARENDRA S.
; TITLE OF INVENTION: ACYL-ACP THIOESTERASES GENES
; TITLE OF INVENTION: AND THEIR USE IN ALTERING PLANT
; TITLE OF INVENTION: OIL COMPOSITION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,176
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/631,264

;; FILING DATE: DECEMBER 20, 1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: CHRISTENBURY, LYNNE M.
;; REGISTRATION NUMBER: 30,971
;; REFERENCE/DOCKET NUMBER: CR-8926-C
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 302-992-5481
;; TELEFAX: 302-773-0164
;; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 362 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-948-176-25

Alignment Scores: 68.8 Length: 362
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-948-176-25 (1-362)

QY 1382 CTCCTCTCTCTAAGTTGAGTGCT 1405
|||||
Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 9

US-08-440-845D-6
; Sequence 6, Application US/08440845D
; Patent No. 5955329
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; APPLICANT: Kridl, J.
; APPLICANT: Dehesh, K.
; APPLICANT: Knauf, V.
; TITLE OF INVENTION: Engineering Plant Thioesterases For
; TITLE OF INVENTION: Altered Substrate Specificity.
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,845D
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 113
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 residues
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-440-845D-6

Alignment Scores: 68.8 Length: 362
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-440-845D-6 (1-362)

QY 1382 CTCCTCTCTCTAAGTTGAGTGCT 1405
|||||
Db 336 LeuHisLeuLeuArgLeuSerGly 343

RESULT 10

US-08-868-458-6
; Sequence 6, Application US/08868458
; Patent No. 6150512
; GENERAL INFORMATION:
; APPLICANT: Yuan, L.
; TITLE OF INVENTION: Engineering Plant Thioesterases And
; Disclosure of Plant Thioesterases
; Having No. 6150512el Substrate Specificity
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
; COMPUTER: IBM PC
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: Microsoft Word For Window 95 7.0a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868,458
; FILING DATE: 03-Jun-1997
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/07064
; FILING DATE: 15-MAY-96
; APPLICATION NUMBER: 08/537,083
; FILING DATE: 29-SEPT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 133
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 residues
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-08-868-458-6

Alignment Scores: 68.8 Length: 362
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-08-868-458-6 (1-362)

.....

; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-459-133-13

Alignment Scores:
Pred. No.: 68.1 Length: 389
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.63%
Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-459-133-13 (1-389)

QY 1568 GTCGCTATTGTGCTGCAGTA 1545
|||||
Db 139 ValProTyrLeuLeuAlaVal 146

RESULT 15

US-09-108-020-53
; Sequence 53, Application US/09108020A
; Patent No. 6143561
; GENERAL INFORMATION:
; APPLICANT: Randall, Douglas D.
; APPLICANT: Johnston, Mark L.
; APPLICANT: Miernyk, Jan A.
; APPLICANT: Luethy, Michael H.
; APPLICANT: Mooney, Brian P.
; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
; FILE REFERENCE: UMO 1482
; CURRENT APPLICATION NUMBER: US/09/108,020A
; CURRENT FILING DATE: 1998-06-30
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/051,291
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/055,255
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/076,544
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Bovine
US-09-108-020-53

Alignment Scores:
Pred. No.: 68 Length: 391
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.63%
Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-108-020-53 (1-391)

QY 2289 GAGCGTTCTTGTCTACTGTCCT 2266
|||||
Db 200 GluAlaPheAlaHisCysPro 207

Search completed: November 30, 2002, 18:51:23
Job time : 49 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:48:35 ; Search time 16.5 Seconds

(without alignments)

7649.408 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atgcaagacgacagcataga.....tggagagtgatgaactttaa 3963

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0

Ygapop 60.0 , Ygapext 60.0

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 102317 seqs, 15924203 residues

Word size: 1

Total number of hits satisfying chosen parameters: 183048

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters: -MODEL=frame+n2p.model -DEV=xlh

-Q-/cgn2_1/USFTO_spool/US09818990/runat_26112002_093224_22184/app_query.fasta_1.4103
-DB-PublishedApplications_AA -QMT-fastan -SUFFIX=oligo.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCLIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09818990.cgn_1.1.10_arunat_26112002_093224_22184
-NCPU=6 -ICPU=3 -NO_XLPU -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261	19.8	264	10	US-09-764-853-638
2	113	8.6	159	10	US-09-764-853-817
3	19	1.4	544	10	US-09-925-299-1004
4	8	0.6	17	10	US-09-864-761-45408
					Sequence 638, App
					Sequence 817, App
					Sequence 1004, Ap
					Sequence 45408, A

46	10	US-09-864-761-38762	0.6	8	Sequence 38762, A
51	10	US-09-864-761-39357	0.6	8	Sequence 39357, A
73	10	US-09-867-550-1662	0.6	8	Sequence 1662, Ap
94	10	US-09-858-664A-23	0.6	8	Sequence 23, Appl
119	10	US-09-858-664A-30	0.6	8	Sequence 30, Appl
257	10	US-09-925-302-757	0.6	8	Sequence 757, App
339	10	US-09-925-297-706	0.6	8	Sequence 706, App
352	10	US-09-134-262-2	0.6	8	Sequence 2, Appli
378	10	US-09-801-368-330	0.6	8	Sequence 330, App
595	10	US-09-815-242-12107	0.6	8	Sequence 12107, A
646	10	US-09-777-710A-1	0.6	8	Sequence 1, Appli
670	10	US-09-777-710A-15	0.6	8	Sequence 15, Appl
746	9	US-10-106-092-2	0.6	8	Sequence 2, Appli
848	10	US-09-839-185-8	0.6	8	Sequence 8, Appli
1286	9	US-10-017-216-7	0.6	8	Sequence 7, Appli
1597	9	US-10-017-216-6	0.6	8	Sequence 6, Appli
1641	9	US-10-017-216-5	0.6	8	Sequence 5, Appli
2053	9	US-10-017-216-2	0.6	8	Sequence 2, Appli
2054	12	US-10-028-946-2	0.6	8	Sequence 2, Appli
2055	9	US-10-017-216-4	0.6	8	Sequence 4, Appli
3594	10	US-09-911-842-4	0.6	8	Sequence 4, Appli
10	10	US-09-884-441-427	0.5	7	Sequence 427, App
11	8	US-08-424-550B-521	0.5	7	Sequence 521, App
21	10	US-09-884-441-402	0.5	7	Sequence 402, App
24	10	US-09-864-761-44088	0.5	7	Sequence 44088, A
25	10	US-09-864-761-44777	0.5	7	Sequence 44777, A
28	10	US-09-982-172-226	0.5	7	Sequence 226, App
30	10	US-09-864-761-47695	0.5	7	Sequence 47695, A
34	12	US-10-001-870-183	0.5	7	Sequence 183, App
37	10	US-09-982-172-94	0.5	7	Sequence 94, Appl
39	10	US-09-864-761-42302	0.5	7	Sequence 42302, A
41	9	US-09-989-919-120	0.5	7	Sequence 120, App
42	10	US-09-726-643-147	0.5	7	Sequence 147, App
42	10	US-09-864-761-45179	0.5	7	Sequence 45179, A
47	10	US-09-864-761-44811	0.5	7	Sequence 44811, A
48	10	US-09-864-761-41726	0.5	7	Sequence 41726, A
51	10	US-09-864-761-47816	0.5	7	Sequence 47816, A
52	10	US-09-864-761-35437	0.5	7	Sequence 35437, A
53	10	US-09-864-761-35128	0.5	7	Sequence 35128, A
61	10	US-09-864-761-41115	0.5	7	Sequence 41115, A
68	10	US-09-925-300-1560	0.5	7	Sequence 1560, Ap
72	10	US-09-891-171-11	0.5	7	Sequence 11, Appl
74	10	US-09-764-853-448	0.5	7	Sequence 448, App
78	9	US-09-749-637A-399	0.5	7	Sequence 399, App
78	10	US-09-925-297-748	0.5	7	Sequence 748, App
89	10	US-09-864-761-43218	0.5	7	Sequence 43218, A
93	10	US-09-864-761-38395	0.5	7	Sequence 38395, A
95	10	US-09-841-879B-12	0.5	7	Sequence 12, Appl
98	9	US-10-103-511-8	0.5	7	Sequence 8, Appli
98	10	US-09-867-550-442	0.5	7	Sequence 442, App
98	10	US-09-731-872-343	0.5	7	Sequence 343, App
98	10	US-09-805-204-8	0.5	7	Sequence 8, Appli
101	9	US-09-286-488-68	0.5	7	Sequence 68, Appl
101	10	US-09-737-178-68	0.5	7	Sequence 68, Appl
105	9	US-09-286-488-69	0.5	7	Sequence 69, Appl
105	10	US-09-737-178-69	0.5	7	Sequence 69, Appl
110	10	US-09-765-272-102	0.5	7	Sequence 102, App
110	10	US-09-800-723-108	0.5	7	Sequence 108, App
112	10	US-09-891-171-7	0.5	7	Sequence 7, Appli
116	9	US-09-896-738-7	0.5	7	Sequence 7, Appli
116	10	US-09-850-165-82	0.5	7	Sequence 82, Appl
116	10	US-09-850-165-111	0.5	7	Sequence 111, App
119	10	US-09-867-550-246	0.5	7	Sequence 246, App
119	10	US-09-731-872-319	0.5	7	Sequence 319, App
120	10	US-09-764-877-1102	0.5	7	Sequence 1102, Ap
121	9	US-09-852-797-68	0.5	7	Sequence 68, Appl
121	9	US-09-852-797-85	0.5	7	Sequence 85, Appl
121	10	US-09-853-161-68	0.5	7	Sequence 68, Appl
121	10	US-09-853-161-85	0.5	7	Sequence 85, Appl
121	10	US-09-852-659A-68	0.5	7	Sequence 68, Appl
121	10	US-09-852-659A-85	0.5	7	Sequence 85, Appl
121	12	US-10-115-406-18	0.5	7	Sequence 18, Appl

c 78	Sequence 44, Appl	c 151	7	0.5	279	10	US-09-741-152-4	Sequence 4, Appl
c 79	Sequence 22, Appl	c 152	7	0.5	279	10	US-09-874-585B-34	Sequence 34, Appl
c 80	Sequence 18, Appl	c 153	7	0.5	280	10	US-09-925-300-1411	Sequence 1411, Ap
c 81	Sequence 997, App	c 154	7	0.5	280	12	US-10-001-843-145	Sequence 145, App
c 82	Sequence 201, Appl	c 155	7	0.5	282	9	US-09-992-598-291	Sequence 291, App
c 83	Sequence 31, Appl	c 156	7	0.5	282	9	US-09-896-738-2	Sequence 2, Appl
c 84	Sequence 31, Appl	c 157	7	0.5	282	9	US-09-915-789A-5	Sequence 5, Appl
c 85	Sequence 359, App	c 158	7	0.5	282	10	US-09-778-320-208	Sequence 208, App
c 86	Sequence 359, App	c 159	7	0.5	282	10	US-09-850-178-33	Sequence 33, Appl
c 87	Sequence 359, App	c 160	7	0.5	282	10	US-09-877-065-8	Sequence 8, Appl
c 88	Sequence 359, App	c 161	7	0.5	282	10	US-09-989-722-291	Sequence 291, App
c 89	Sequence 359, App	c 162	7	0.5	282	10	US-09-989-723-291	Sequence 291, App
c 90	Sequence 359, App	c 163	7	0.5	282	10	US-09-989-729-291	Sequence 291, App
c 91	Sequence 359, App	c 164	7	0.5	282	10	US-09-989-727-291	Sequence 291, App
c 92	Sequence 359, App	c 165	7	0.5	282	10	US-09-910-689-208	Sequence 208, App
c 93	Sequence 359, App	c 166	7	0.5	282	10	US-09-989-731-291	Sequence 291, App
c 94	Sequence 359, App	c 167	7	0.5	282	10	US-09-884-441-333	Sequence 333, App
c 95	Sequence 359, App	c 168	7	0.5	282	10	US-09-989-732-291	Sequence 291, App
c 96	Sequence 359, App	c 169	7	0.5	282	10	US-09-991-073-291	Sequence 291, App
c 97	Sequence 359, App	c 170	7	0.5	282	10	US-09-990-442-291	Sequence 291, App
c 98	Sequence 444, App	c 171	7	0.5	282	10	US-09-993-604-291	Sequence 291, App
c 99	Sequence 431, App	c 172	7	0.5	282	10	US-09-993-604-291	Sequence 291, App
c 100	Sequence 4, Appl	c 173	7	0.5	282	10	US-09-990-456-291	Sequence 291, App
c 101	Sequence 26, Appl	c 174	7	0.5	282	10	US-09-989-721-291	Sequence 291, App
c 102	Sequence 164, App	c 175	7	0.5	282	12	US-10-006-867-60	Sequence 60, Appl
c 103	Sequence 77, Appl	c 176	7	0.5	282	12	US-10-052-586-218	Sequence 218, App
c 104	Sequence 1154, Ap	c 177	7	0.5	282	12	US-10-010-742-208	Sequence 208, App
c 105	Sequence 78, Appl	c 178	7	0.5	306	9	US-09-712-363-186	Sequence 186, App
c 106	Sequence 183, App	c 179	7	0.5	307	9	US-09-764-868-1100	Sequence 1100, Ap
c 107	Sequence 34202, A	c 180	7	0.5	309	10	US-09-778-320-209	Sequence 209, App
c 108	Sequence 34988, A	c 181	7	0.5	309	10	US-09-910-689-209	Sequence 209, App
c 109	Sequence 36844, A	c 182	7	0.5	309	10	US-09-884-441-332	Sequence 392, App
c 110	Sequence 731, App	c 183	7	0.5	309	12	US-10-010-742-209	Sequence 209, App
c 111	Sequence 39, Appl	c 184	7	0.5	310	9	US-09-286-488-19	Sequence 19, Appl
c 112	Sequence 1104, Ap	c 185	7	0.5	310	10	US-09-737-178-19	Sequence 19, Appl
c 113	Sequence 101, App	c 186	7	0.5	310	10	US-09-925-300-1602	Sequence 1602, App
c 114	Sequence 2, Appl	c 187	7	0.5	312	10	US-09-925-302-783	Sequence 783, App
c 115	Sequence 19, Appl	c 188	7	0.5	314	10	US-09-768-826-43	Sequence 43, Appl
c 116	Sequence 10, Appl	c 189	7	0.5	315	10	US-09-815-242-11636	Sequence 11636, A
c 117	Sequence 20, Appl	c 190	7	0.5	319	10	US-09-925-302-758	Sequence 758, App
c 118	Sequence 10825, A	c 191	7	0.5	319	10	US-09-886-055-95	Sequence 95, Appl
c 119	Sequence 677, App	c 192	7	0.5	325	10	US-09-927-113-10	Sequence 10, Appl
c 120	Sequence 18, Appl	c 193	7	0.5	331	9	US-09-941-831-17	Sequence 17, Appl
c 121	Sequence 11, Appl	c 194	7	0.5	338	10	US-09-741-669-365	Sequence 365, App
c 122	Sequence 817, App	c 195	7	0.5	338	10	US-09-912-020-373	Sequence 373, App
c 123	Sequence 5404, Ap	c 196	7	0.5	340	10	US-09-815-242-13257	Sequence 13257, A
c 124	Sequence 12578, A	c 197	7	0.5	340	10	US-09-815-242-13579	Sequence 13579, A
c 125	Sequence 12778, A	c 198	7	0.5	349	10	US-09-919-497-76	Sequence 76, Appl
c 126	Sequence 54, Appl	c 199	7	0.5	352	9	US-09-104-063-4	Sequence 4, Appl
c 127	Sequence 68, Appl	c 200	7	0.5	352	10	US-09-953-692-2	Sequence 2, Appl
c 128	Sequence 74, Appl	c 201	7	0.5	352	10	US-09-953-717-2	Sequence 2, Appl
c 129	Sequence 488, App	c 202	7	0.5	354	10	US-09-925-302-501	Sequence 501, App
c 130	Sequence 1616, Ap	c 203	7	0.5	356	10	US-09-925-300-1326	Sequence 1326, Ap
c 131	Sequence 19, Appl	c 204	7	0.5	367	9	US-09-286-488-20	Sequence 20, Appl
c 132	Sequence 1678, Ap	c 205	7	0.5	367	9	US-09-813-398-18	Sequence 18, Appl
c 133	Sequence 95, Appl	c 206	7	0.5	367	10	US-09-737-178-20	Sequence 20, Appl
c 134	Sequence 10415, A	c 207	7	0.5	369	9	US-09-895-913A-180	Sequence 180, App
c 135	Sequence 3, Appl	c 208	7	0.5	371	10	US-09-764-853-600	Sequence 600, App
c 136	Sequence 17, Appl	c 209	7	0.5	371	10	US-09-764-853-790	Sequence 790, App
c 137	Sequence 27, Appl	c 210	7	0.5	372	10	US-09-768-894A-2	Sequence 2, Appl
c 138	Sequence 18, Appl	c 211	7	0.5	380	10	US-09-815-242-11249	Sequence 11249, A
c 139	Sequence 143, App	c 212	7	0.5	414	9	US-10-108-605-87	Sequence 1027, Ap
c 140	Sequence 183, App	c 213	7	0.5	415	9	US-09-764-868-1027	Sequence 87, Appl
c 141	Sequence 446, App	c 214	7	0.5	427	9	US-09-975-253-2	Sequence 2, Appl
c 142	Sequence 3, Appl	c 215	7	0.5	433	10	US-09-801-368-38	Sequence 38, Appl
c 143	Sequence 12, Appl	c 216	7	0.5	434	10	US-09-801-574-78	Sequence 78, Appl
c 144	Sequence 2, Appl	c 217	7	0.5	448	10	US-09-815-242-12017	Sequence 12017, A
c 145	Sequence 2, Appl	c 218	7	0.5	450	10	US-09-765-111A-14	Sequence 14, Appl
c 146	Sequence 2, Appl	c 219	7	0.5	453	10	US-09-815-242-10872	Sequence 10872, A
c 147	Sequence 5, Appl	c 220	7	0.5	458	10	US-09-826-463-1	GENERAL INFORMA
c 148	Sequence 14, Appl	c 221	7	0.5	459	10	US-09-771-161A-198	Sequence 198, App
c 149	Sequence 640, App	c 222	7	0.5	474	9	US-09-998-909-14	Sequence 14, Appl
c 150	Sequence 34, Appl	c 223	7	0.5	476	10	US-09-925-300-1582	Sequence 1582, Ap

c 954	6	0.5	126	10	US-09-730-857-73	Sequence 73, Appl
c 955	6	0.5	126	10	US-09-730-857-77	Sequence 77, Appl
c 956	6	0.5	127	10	US-09-925-302-882	Sequence 882, App
c 957	6	0.5	127	10	US-09-864-761-35071	Sequence 35071, A
c 958	6	0.5	127	10	US-09-809-739-10	Sequence 10, Appl
c 959	6	0.5	127	10	US-09-998-831-7	Sequence 7, Appl
c 960	6	0.5	128	9	US-09-989-920-220	Sequence 220, App
c 961	6	0.5	128	10	US-09-864-761-44943	Sequence 44943, A
c 962	6	0.5	128	10	US-09-925-299-793	Sequence 793, App
c 963	6	0.5	128	10	US-09-925-299-898	Sequence 898, App
c 964	6	0.5	128	10	US-09-925-297-603	Sequence 603, App
c 965	6	0.5	129	10	US-09-925-300-1573	Sequence 1573, Ap
c 966	6	0.5	129	10	US-09-864-761-40670	Sequence 40670, A
c 967	6	0.5	129	10	US-09-764-869-633	Sequence 633, App
c 968	6	0.5	129	10	US-09-965-099-99	Sequence 99, Appl
c 969	6	0.5	129	12	US-10-051-852-99	Sequence 99, Appl
c 970	6	0.5	130	10	US-09-815-242-4999	Sequence 4999, Ap
c 971	6	0.5	130	10	US-09-916-790-13	Sequence 13, Appl
c 972	6	0.5	130	10	US-09-730-617-12	Sequence 12, Appl
c 973	6	0.5	130	10	US-09-730-617-85	Sequence 85, Appl
c 974	6	0.5	130	10	US-09-730-617-87	Sequence 87, Appl
c 975	6	0.5	130	10	US-09-730-617-88	Sequence 88, Appl
c 976	6	0.5	130	10	US-09-730-617-97	Sequence 97, Appl
c 977	6	0.5	130	12	US-10-071-751-39	Sequence 39, Appl
c 978	6	0.5	131	9	US-09-982-992A-2	Sequence 2, Appli
c 979	6	0.5	131	9	US-09-984-245-145	Sequence 145, App
c 980	6	0.5	131	10	US-09-564-329A-5	Sequence 5, Appli
c 981	6	0.5	131	10	US-09-864-761-42682	Sequence 42682, A
c 982	6	0.5	131	10	US-09-452-239-30	Sequence 30, Appl
c 983	6	0.5	131	10	US-09-855-153-5	Sequence 5, Appli
c 984	6	0.5	131	10	US-09-854-811-5	Sequence 5, Appli
c 985	6	0.5	131	10	US-09-934-773-5	Sequence 5, Appli
c 986	6	0.5	131	10	US-09-963-620-5	Sequence 5, Appli
c 987	6	0.5	131	12	US-10-101-747-2	Sequence 2, Appli
c 988	6	0.5	132	10	US-09-759-143-573	Sequence 573, App
c 989	6	0.5	132	10	US-09-764-870-332	Sequence 332, App
c 990	6	0.5	132	10	US-09-780-669-573	Sequence 573, App
c 991	6	0.5	132	10	US-09-925-297-657	Sequence 657, App
c 992	6	0.5	132	10	US-09-822-827-573	Sequence 573, App
c 993	6	0.5	132	10	US-09-764-903-61	Sequence 61, Appl
c 994	6	0.5	132	10	US-09-771-161A-158	Sequence 158, App
c 995	6	0.5	132	10	US-09-922-261-54	Sequence 54, Appl
c 996	6	0.5	132	12	US-10-078-929-32	Sequence 32, Appl
c 997	6	0.5	134	9	US-10-051-325-6	Sequence 6, Appli
c 998	6	0.5	135	10	US-09-788-074-1	Sequence 1, Appli
c 999	6	0.5	135	10	US-09-759-143-884	Sequence 884, App
c1000	6	0.5	135	10	US-09-855-271-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-764-853-638
; Sequence 638, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: Puz06
; CURRENT APPLICATION NUMBER: US/09764, 853
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 638
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-853-638

Alignment Scores: 2.46e-233 Length: 264
Pred. No.: 261.00 Matches: 261
Score:

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.83% Indels: 0
DB: 10 Gaps: 0
US-09-818-990b-1 (1-3963) x US-09-764-853-638 (1-264)
QY 3178 GAAAGACAAAGAGCCCTACAGAAAGCGCTTTTTCGACACACATTTCCTCGACGGCTCCT 3237
Db 4 GluArgAspLysGluProLeuGlnGluArgPheArgProHisPheLeuGlnAlaPro 23
QY 3238 GGGGATATGGTAGCTCATGAGGGCGCCTCTGTCGGCTGGACTCTAAAGGTGAGTGGTTTA 3297
Db 24 GlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLeu 43
QY 3298 CCGCCCCGGAGCTGACATGGCTACTCAATGGCCAACTGTGCTACACAGATGCCTCCAC 3357
Db 44 ProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHis 63
QY 3358 AAGATGCTGGTCAGGAGACCGGAGTCCACTCTCTGCTCATTGACCCACTCCTCAGCGC 3417
Db 64 LysMetLeuValArgGluThrGlyValHisSerLeuLeuLeuAspProLeuThrGlnArg 83
QY 3418 GACGACGGACCTATAAGTGCATCGCTACCAACAAAAACCGGCGAGAAATCTTTTAGTCTG 3477
Db 84 AspAlaGlyThrTyLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeu 103
QY 3478 GAGCTCTCTGTAGTAGCAAGAGGTGAAGAAACACTGTGATCCTGGGAGAACTACAG 3537
Db 104 GluLeuSerValValAlaLysGluValLysAlaProValIleLeuGluLysLeuGln 123
QY 3538 AACTGCGGTGTTCCGAGGCCACCCGTCGAGAGCTGAGTCCGCGCTGATAGGATGCC 3597
Db 124 AsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPro 143
QY 3598 CCACCTGTGTTCTACTGGAAGAAAGACAATGAGACCATCCCTTGCACAGAGAGGATC 3657
Db 144 ProProValPheTyTrpLysLysAspAsnGluThrIleProCysThrArgGluArgile 163
QY 3658 AGTATGACACAGGACACAAACAGGTATGCCCTTCTCATTCAGCCAGCCCAAAATCA 3717
Db 164 SerMetHisGlnAspThrThrGlyTyAlaCysLeuLeuGlnProAlaLysLysSer 183
QY 3718 GACGCTGGATGGTACAGCTTGTACGCAAGAAATCAAGCCGCGCATCGTGTGCTGACATGCC 3777
Db 184 AsplaglyTrpTyThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAla 203
QY 3778 AGGTGGATATATACGCTCAGTGGCACCATCAGATCCACGCCCATGTCTGTCCGGGCC 3837
Db 204 ArgLeuAspIleTyAlaGlnTrpHisGlnIleProProMetSerValArgPro 223
QY 3838 AGTGGCAGTCGCTACGATCTCTCACCAGTAAAGGACTTGACATATTTTTCGCTTTTCC 3897
Db 224 SerGlySerArgTyGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSer 243
QY 3898 TCCATGGAAGACGATGGTGTATTTCATGCTCTCTCGGAGTGTAGTGGAGAGTGTAGTAA 3957
Db 244 SerMetGluSerThrMetValTySerCysSerSerArgSerValValGluSerAspGlu 263
QY 3958 CTT 3960
Db 264 Leu 264
RESULT 2
US-09-764-853-817
; Sequence 817, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: Puz06
; CURRENT APPLICATION NUMBER: US/09764, 853
; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 817
; LENGTH: 159

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (121)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (139)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-817

Alignment Scores:
Pred. No.: 2,738-96 Length: 159
Score: 113.00 Matches: 113
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.59% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-764-853-817 (1-159)

QY 3424 GGCACCTAATAGTCATCGTACCAACAAACCGGCGACAATCTTTTAGCTGGAGCTC 3483
Db 8 GlyThrTyrLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLeuGluLeu 27

QY 3484 TCTGTAGTACCCAAAGAGGTGAAGAACGACCTGTGATCTCTGGAGAACTACAGAACTGC 3543
Db 28 SerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGlnAsnCys 47

QY 3544 GGTGTTCCGAGGCCACCCGTGAGACTGGAGTCCGGGTGATAGCATGCCCCACCT 3603
Db 48 GlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetProPro 67

QY 3604 GTGTTCTACTGGAAGAAACATGAGACCATCCCTTGACACAGAGAGGATCAGTATG 3663
Db 68 ValPheTyrTrpLysLysAspLsnGluThrIleProCysThrArgGluArgIleSerMet 87

QY 3664 CACGAGACACACAGGGTATGCTCCCTTCTCATTTACAGCCAGCCAAAGAAATCAGACGCT 3723
Db 88 HisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSerAspAla 107

QY 3724 GGATGTACACGTGTCAGCCAGATGAAGCCGGCATC 3762
Db 108 GlyTrpTyrThrLeuSerAlaLysAsnGluAlaGlyIle 120

RESULT 3

US-09-925-299-1004
; Sequence 1004, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1004

; LENGTH: 544
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (531)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1004

Alignment Scores:
Pred. No.: 2,528-09 Length: 544
Score: 19,00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-925-299-1004 (1-544)

QY 3034 GATGACGATGGCAACTACACCATCATGGCAGCAACCCCGGAGAGATCAGCTGT 3090
Db 229 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 247

RESULT 4

US-09-864-761-45408
; Sequence 45408, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408

```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45408
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL137225.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.75
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
US-09-864-761-45408

Alignment Scores:
Pred. No.: 57.4 Length: 17
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990b-1 (1-3963) x US-09-864-761-45408 (1-17)

QY 1185 TCCAGCAGTACCCAGCCAGCA 1208
|||||
Db 2 SerSerSerThrProSerProAla 9

RESULT 5
US-09-864-761-38762
; Sequence 38762, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomlca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38762
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002472.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
; OTHER INFORMATION: SWISSPROT HIT: P03179, EVALUE 4.80e+00
US-09-864-761-38762

Alignment Scores:
Pred. No.: 51.2 Length: 46
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990b-1 (1-3963) x US-09-864-761-38762 (1-46)

QY 3146 CCAGCAGAGGTAGCCGACTGCGA 3123
|||||
Db 35 ProAlaGluValSerArgLeuArg 42

RESULT 6
US-09-864-761-39357
; Sequence 39357, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomlca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669

```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39357
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004590.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: SWISSPROT HIT: Q15438, EVALUE 4.00e-10
; OTHER INFORMATION: EST_HUMAN HIT: BF690441.1, EVALUE 2.90e+00
US-09-864-761-39357

Alignment Scores:
Pred. No.: 50.6 Length: 51
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-864-761-39357 (1-51)
QY 196 TTTCTGAGCCCAAGAAATTAGAC 219
Db 11 PheLeuSerGlnGluLeuAsp 18

RESULT 7
US-09-867-550-1662
; Sequence 1662, Application US/09867550
; Patent No. US20020082206N1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1662
```

```
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-1662

Alignment Scores:
Pred. No.: 48.5 Length: 73
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-867-550-1662 (1-73)
QY 111 CTCCTCAACCCCTTGCATTTCCG 134
Db 2 LeuLeuGlnProLeuProPheArg 9

RESULT 8
US-09-858-664A-23
; Sequence 23, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-23

Alignment Scores:
Pred. No.: 47.1 Length: 94
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-858-664A-23 (1-94)
QY 3313 ACATGGCTACTCAATGGCCAACT 3336
Db 34 ThrTriLeuLeuAsnGlyGlnPro 41

RESULT 9
US-09-858-664A-30
; Sequence 30, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 119
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-30

Alignment Scores:
Pred. No.: 45.8 Length: 119
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-858-664A-30 (1-119)
QY 3313 ACATGGCTACTCAATGCGCCAACT 3336
|||||
Db 57 ThrTrpLeuLeuAsnGlyGlnPro 64

RESULT 10
US-09-925-302-757
; Sequence 757, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 757
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (210)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-757

Alignment Scores:
Pred. No.: 41.9 Length: 257
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-925-302-757 (1-257)
QY 174 ATCTGGCCTCCGCTCCTTCAGC 151
|||||
Db 90 IleLeuAlaSerAlaSerPheSer 97

RESULT 11
US-09-925-297-706
; Sequence 706, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 706
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (173)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-706

Alignment Scores:
Pred. No.: 40.6 Length: 339
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-925-297-706 (1-339)
QY 196 TTTCTGAGCCCAAGAATAATTAGAC 219
|||||
Db 310 PheLeuSerGlnGluLeuAsp 317

RESULT 12
US-09-134-262-2
; Sequence 2, Application US/09134262
; Patent No. US20010002489A1
; GENERAL INFORMATION:
; APPLICANT: Calgene LLC
; TITLE OF INVENTION: Methods for Increasing Stearate Content in Soybean Oil
; FILE REFERENCE: Docket #17030
; CURRENT APPLICATION NUMBER: US/09/134,262
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: IBM PC; Windows NT 4.0; Microsoft Word For Windows 7.0a
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Garcinia mangostana
US-09-134-262-2

Alignment Scores:
Pred. No.: 40.4 Length: 352
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-134-262-2 (1-352)
QY 1382 CTCATCTCCTAAGGTTGAGTGT 1405
|||||
Db 327 LeuHisLeuLeuArgLeuSerGly 334

RESULT 13
US-09-801-368-330
; Sequence 330, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
```

```

; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 330
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-330

Alignment Scores:
Pred. No.: 40.1 Length: 378
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-801-368-330 (1-378)
QY 2795 ATTTCATCATCTGATTCATCAACA 2772
|||||
Db 118 lIeSerSerAspSerSerThr 125

RESULT 14
US-09-815-242-12107
; Sequence 12107, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12107
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12107

Alignment Scores:
Pred. No.: 38 Length: 595
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-815-242-12107 (1-595)
QY 2371 CTGTTGGCCTGGTGGAGTGGCT 2348
|||||
Db 174 LeuLeuGlyLeuValGlyValAla 181

RESULT 15
US-09-777-710A-1
; Sequence 1, Application US/09777710A
; Patent No. US20020058305A1
; GENERAL INFORMATION:
; APPLICANT: OKINO, No. US20020058305A1omu et al.
; TITLE OF INVENTION: CERAMIDASE GENE
; FILE REFERENCE: 1422-0458P
; CURRENT APPLICATION NUMBER: US/09/777,710A
; CURRENT FILING DATE: 2001-02-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-777-710A-1

Alignment Scores:
Pred. No.: 37.6 Length: 646
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x US-09-777-710A-1 (1-646)
QY 718 CAGCTGCCAGTGAGCGGCTGGT 741
|||||
Db 425 GluAlaAlaSerGluAlaAlaGly 432

Search completed: November 30, 2002, 19:03:49
Job time : 38.5 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:38:34 ; Search time 52.5 Seconds
(without alignments)
14513.551 Million cell updates/sec

Title: US-09-818-990B-1
Perfect score: 1316
Sequence: 1 atgcaagacgacagcataga.....tggagagtgatgaactttaa 3963

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODE=framer_n2p.model -DEV-xmlh
-Q/cgn2_1/USPTO_spool/US09818990/runat_26112002_093222_22067/app_query.fasta_1.4103
-DB=PIR_73 -QFMT=fastan -SUFFIX=oligo.rpr -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000
-DOCALIGN=200 -THR.SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990_@CGN_1.1.65_@runat_26112002_093222_22067 -NCPU=6 -ICPU=3
-NO_XLIPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -XGAPOP=6 -XGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

		SUMMARIES			
Result No.	Score	Query Match	Length	ID	Description
1	19	1.4	772	T13078	KIAA0992 protein -
C 2	9	0.7	156	2	hypothetical prote
3	9	0.7	393	1	Ig mu chain C regi
4	9	0.7	438	1	Ig mu chain C regi
5	9	0.7	438	1	Ig mu chain C regi
6	9	0.7	461	1	Ig mu chain C regi
C 7	9	0.7	490	2	hypothetical prote
8	9	0.7	587	2	Ca2+/calmodulin-de
C 9	9	0.7	897	2	hypothetical prote
10	9	0.7	1756	2	Tys protein - yeas
11	8	0.6	102	2	hypothetical prote
C 12	8	0.6	107	2	conserved hypothet
13	8	0.6	109	2	hypothetical prote
14	8	0.6	122	2	hypothetical prote

hypothetical prote	122	0.6	D71106
UL33 protein - hum	130	0.6	1 WMBEH3
heat-stable antige	141	0.6	2 S15785
pathogenesis-relat	164	0.6	2 S37166
probable transport	192	0.6	2 H91103
hypothetical prote	20	0.6	2 C85949
probable membrane	21	0.6	2 AG0875
hypothetical prote	192	0.6	2 AF1318
hypothetical prote	198	0.6	2 AF1690
hypothetical prote	199	0.6	2 E85537
hypothetical prote	25	0.6	2 A90687
transcription fact	26	0.6	2 S60325
conserved hypothet	204	0.6	2 F87295
hypothetical prote	228	0.6	2 F90130
conserved hypothet	239	0.6	2 B82361
conserved hypothet	260	0.6	2 A75572
hypothetical prote	272	0.6	2 S24375
hypothetical prote	273	0.6	2 A11989
probable enoyl-coA	274	0.6	2 B70695
cell-shape determi	295	0.6	2 AC1268
alsSD operon activ	302	0.6	2 A47126
probable coaA prot	312	0.6	2 B70896
hypothetical prote	313	0.6	2 B96692
iron(III) ABC tran	322	0.6	2 G72421
homeotic protein H	323	0.6	2 S16318
hypothetical prote	356	0.6	2 E84199
hypothetical prote	360	0.6	2 H95172
DEAD RNA helicase	360	0.6	2 G98038
probable oleoyl-carri	362	0.6	2 T14393
3-methyl-2-oxobuta	369	0.6	1 S28950
probable ABC trans	369	0.6	2 D81261
pectate lyase (EC	374	0.6	2 JC1313
pectase lyase 2 pr	374	0.6	2 S49306
SLG1 protein - yea	378	0.6	2 S61992
hypothetical prote	381	0.6	2 C90543
hypothetical prote	382	0.6	2 S40987
benzoate membrane	383	0.6	2 AF3503
3-methyl-2-oxobuta	390	0.6	2 S39807
molybdopterin bios	390	0.6	2 S75715
3-methyl-2-oxobuta	392	0.6	1 A34267
3-methyl-2-oxobuta	392	0.6	1 A37157
carbamoylphosphate	401	0.6	2 AI2842
threonine synthase	415	0.6	2 F84393
uroporphyrin-III C	424	0.6	2 G81160
hypothetical prote	424	0.6	2 D81946
chemotaxis MotB pr	427	0.6	2 AC0203
sugar kinase (impo	428	0.6	2 AC2986
hypothetical prote	456	0.6	2 B97620
probable rhamnose	460	0.6	2 D98297
atrazine chlorohyd	463	0.6	2 T38111
probable transport	478	0.6	2 A82336
catalase (EC 1.1.1.	487	0.6	2 T34858
alpha-interneixin -	504	0.6	2 I53868
probable sugar tra	511	0.6	2 H84536
cellulose 1,4-beta	511	0.6	2 A84537
cellulose bZIP tran	516	0.6	2 JS0083
hypothetical prote	516	0.6	2 S33164
hypothetical prote	519	0.6	2 G84598
hypothetical prote	528	0.6	2 B84743
hypothetical prote	531	0.6	2 T02264
hypothetical prote	557	0.6	2 S62002
hypothetical prote	559	0.6	2 AI2227
probable two-compo	595	0.6	2 C82960
receptor-like prot	624	0.6	2 C96756
hypothetical prote	624	0.6	2 S67382
peptidylprolyl iso	628	0.6	2 AG3357
conserved hypothet	630	0.6	2 T38023
glycyl-tRNA synthe	670	0.6	2 C83540
hypothetical prote	684	0.6	2 D83645
glycyl-tRNA synthe	694	0.6	2 T10565
VacB protein Xf198	717	0.6	2 F82613
hypothetical prote	717	0.6	2 T49238

88	8	0.6	748	2	T04011	hypothetical prote	161	7	0.5	120	2	S50867	4E-BP2 protein - h
89	8	0.6	767	2	T19690	hypothetical prote	c 162	7	0.5	121	2	PT0378	Ig heavy chain v r
90	8	0.6	834	2	R82673	hypothetical prote	c 163	7	0.5	121	2	PT0391	Ig heavy chain v r
91	8	0.6	881	2	T52601	squamosa promoter	c 164	7	0.5	121	2	E30502	Ig heavy chain v r
92	8	0.6	881	2	T52602	squamosa promoter	c 165	7	0.5	121	2	E64391	hypothetical prote
93	8	0.6	925	2	D59105	hypothetical prote	c 166	7	0.5	121	2	G90435	hypothetical prote
94	8	0.6	1051	2	A35761	cell surface glyco	c 167	7	0.5	122	1	AVMSX2	Ig heavy chain v r
95	8	0.6	1095	1	A31225	phospholipase C (E	c 168	7	0.5	122	2	AI3651	hypothetical prote
96	8	0.6	1152	2	H86486	protein Tyl/copa-	c 169	7	0.5	123	2	PT0380	Ig heavy chain v r
97	8	0.6	1337	1	I38670	protein-tyrosine-p	c 170	7	0.5	123	2	PT0381	Ig heavy chain v r
98	8	0.6	1339	2	A84583	probable SNF2 subf	c 171	7	0.5	123	2	PT0383	Ig heavy chain v r
99	8	0.6	1527	2	JE0336	canalicular multib	c 172	7	0.5	123	2	PT0384	Ig heavy chain v r
100	8	0.6	1581	2	B71636	hypothetical prote	c 173	7	0.5	123	2	PT0385	Ig heavy chain v r
101	8	0.6	1583	2	F97846	hypothetical prote	c 174	7	0.5	123	2	PT0386	Ig heavy chain v r
102	8	0.6	1597	2	S68420	citron - mouse	c 175	7	0.5	123	2	T45375	hypothetical prote
103	8	0.6	1894	2	JC4980	plexin 1 precursor	c 176	7	0.5	124	1	E69191	conserved hypotet
104	8	0.6	1905	2	I51553	Plexin - African c	c 177	7	0.5	124	2	S25113	insulin-like growt
105	8	0.6	2023	2	T13154	polycomb protein e	c 178	7	0.5	124	2	PT0388	Ig heavy chain v r
106	8	0.6	2102	2	T15626	hypothetical prote	c 179	7	0.5	124	2	PT0389	Ig heavy chain v r
107	8	0.6	2559	2	T30850	fat facets protein	c 180	7	0.5	124	2	F84561	hypothetical prote
108	8	0.6	2747	2	B49132	fat facets (faf) s	c 181	7	0.5	125	2	A53692	synapsin I - mouse
109	8	0.6	4924	2	T50176	probable peptide s	c 182	7	0.5	125	2	E72716	hypothetical prote
110	7	0.5	41	2	A44536	T-cell receptor al	c 183	7	0.5	126	2	S35887	AL5 protein - indl
111	7	0.5	43	2	I41026	colicin 10 lysis p	c 184	7	0.5	126	2	AD2523	hypothetical prote
112	7	0.5	45	2	S10921	colicin E1 lysis p	c 185	7	0.5	126	2	AI2510	hypothetical prote
113	7	0.5	47	2	FC4178	DNA-directed DNA p	c 186	7	0.5	130	2	E86906	ribosomal protein
114	7	0.5	49	1	ZHECP3	colicin lysis prot	c 187	7	0.5	131	2	B60725	hypothetical prote
115	7	0.5	49	2	H97535	colicin N lysis pr	c 188	7	0.5	131	2	H75316	conserved hypotet
116	7	0.5	52	1	ZHECN4	interleukin-5 rece	c 189	7	0.5	134	2	T13292	repressor protein
117	7	0.5	63	2	A60235	pepi protein - Sta	c 190	7	0.5	135	2	T32385	hypothetical prote
118	7	0.5	69	2	S58357	hypothetical prote	c 191	7	0.5	135	2	T49640	hypothetical prote
119	7	0.5	71	2	B81812	photosystem II pho	c 192	7	0.5	136	2	H81891	hypothetical prote
120	7	0.5	73	1	F2NT0P	hypothetical prote	c 193	7	0.5	137	2	S54219	flagellar basal bo
121	7	0.5	76	2	T29579	GASr1-like protein	c 194	7	0.5	137	2	AD0219	NADH2 dehydrogenas
122	7	0.5	80	2	H96775	hypothetical prote	c 195	7	0.5	137	2	S28247	hypothetical prote
123	7	0.5	80	2	A81361	Ig heavy chain v r	c 196	7	0.5	138	2	E75509	beta/gamma crystal
124	7	0.5	91	2	S24523	hypothetical prote	c 197	7	0.5	138	2	C87376	hypothetical prote
125	7	0.5	92	2	C81022	hypothetical prote	c 198	7	0.5	138	2	H97488	ADP-ribose Pyropho
126	7	0.5	93	2	H36808	antifungal protein	c 199	7	0.5	138	2	AH2706	F6F3.17 protein -
127	7	0.5	94	2	S44064	conserved membrane	c 200	7	0.5	139	2	B86144	proteinase umub (E
128	7	0.5	95	2	B87216	hypothetical prote	c 201	7	0.5	139	2	A36713	UmuD protein (EC 3
129	7	0.5	95	2	B81204	hypothetical prote	c 202	7	0.5	139	2	AF0755	hypothetical prote
130	7	0.5	100	2	F81966	hypothetical prote	c 203	7	0.5	142	2	AG2994	transporter, dne f
131	7	0.5	100	2	S67141	hypothetical prote	c 204	7	0.5	143	2	AC3339	subtilisin inhibit
132	7	0.5	100	2	H72680	Ig heavy chain v r	c 205	7	0.5	144	1	XSSNA	hypothetical prote
133	7	0.5	101	2	PU0002	hypothetical prote	c 206	7	0.5	144	2	C71252	Ig heavy chain pre
134	7	0.5	102	2	H75352	hypothetical prote	c 207	7	0.5	145	2	S03844	conserved hypotet
135	7	0.5	102	2	T45500	Ig heavy chain v r	c 208	7	0.5	146	2	C69983	conserved hypotet
136	7	0.5	105	2	PL0255	hypothetical prote	c 209	7	0.5	146	2	T14667	hypothetical prote
137	7	0.5	106	2	JQ1431	conserved hypotet	c 210	7	0.5	146	2	AI2918	hypothetical prote
138	7	0.5	107	2	B81031	somatotropin, vari	c 211	7	0.5	147	2	T16672	hypothetical prote
139	7	0.5	109	2	A29864	hypothetical prote	c 212	7	0.5	148	2	T02665	probable prolamin
140	7	0.5	110	2	C71160	Ig heavy chain v r	c 213	7	0.5	150	2	H97752	hypothetical prote
141	7	0.5	111	2	PH1028	conserved hypotet	c 214	7	0.5	151	2	G86760	diacylglycerol kin
142	7	0.5	111	2	B75351	Ig heavy chain v r	c 215	7	0.5	151	2	T14948	hypothetical prote
143	7	0.5	113	2	PH1029	photosystem I 8.4K	c 216	7	0.5	151	2	E71547	hypothetical prote
144	7	0.5	113	2	S06684	Ig heavy chain v r	c 217	7	0.5	152	2	A87261	hypothetical prote
145	7	0.5	114	2	PH1027	protein-serine/thr	c 218	7	0.5	152	2	T00877	hypothetical prote
146	7	0.5	114	2	I38224	hypothetical prote	c 219	7	0.5	152	2	C98289	hypothetical prote
147	7	0.5	114	2	T25909	ribosomal protein	c 220	7	0.5	154	2	S73656	MG288 homolog F04_
148	7	0.5	114	2	H89937	Ig lambda chain C	c 221	7	0.5	154	2	E72485	E6 protein - human
149	7	0.5	115	2	S78270	probable Arp synth	c 222	7	0.5	156	1	W6WL41	hypothetical prote
150	7	0.5	116	2	I51338	transcription regu	c 223	7	0.5	156	2	A87366	hypothetical prote
151	7	0.5	117	2	F81025	Ig heavy chain v r	c 224	7	0.5	156	2	S52246	transposable retro
152	7	0.5	117	2	D81969	nicotinic acetylch	c 225	7	0.5	157	2	C81443	molybdenum cofacto
153	7	0.5	117	2	H95130	probable oleoyl-fa	c 226	7	0.5	157	2	C82651	hypothetical prote
154	7	0.5	118	2	S24527	Ig heavy chain v r	c 227	7	0.5	159	2	A36107	ribosomal protein
155	7	0.5	118	2	A32673	hypothetical prote	c 228	7	0.5	159	2	H86174	hypothetical prote
156	7	0.5	118	2	T07177	Ig heavy chain v r	c 229	7	0.5	159	2	A90292	hypothetical prote
157	7	0.5	118	2	S27476	Ig heavy chain v r	c 230	7	0.5	160	2	G71122	hypothetical prote
158	7	0.5	119	2	S24522	Ig heavy chain v r	c 231	7	0.5	160	2	G90367	hypothetical prote
159	7	0.5	119	2	S24497	Ig heavy chain v r	c 232	7	0.5	160	2	E82441	conserved hypotet
160	7	0.5	120	2	PT0393	Ig heavy chain v r	c 233	7	0.5	160	2		

234	7	0.5	161	2	H69184	conserved hypother	c 307	7	0.5	209	2	H69037	conserved hypother
235	7	0.5	161	2	F82637	conserved hypother	308	7	0.5	210	2	C84278	hypothetical prote
236	7	0.5	163	2	S33418	complement C4 prot	309	7	0.5	211	2	C84278	hypothetical prote
237	7	0.5	163	2	F72762	hypothetical prote	c 310	7	0.5	212	2	T43308	transaldolase (EC
238	7	0.5	164	2	A86023	hypothetical prote	c 311	7	0.5	214	2	G64432	acetyltransferase
239	7	0.5	164	2	H91176	Shux-like protein	312	7	0.5	214	2	T10681	hypothetical prote
240	7	0.5	165	2	T45271	probable mini-circ	313	7	0.5	215	2	H75282	conserved hypother
241	7	0.5	165	2	A97105	hypothetical prote	314	7	0.5	217	2	T36951	conserved hypother
242	7	0.5	166	2	H87652	chemotaxis protein	315	7	0.5	217	2	T40730	probable rna-bindi
243	7	0.5	167	2	B98088	conserved hypother	316	7	0.5	218	2	C71031	probable ribosomal
244	7	0.5	168	2	AF2620	conserved hypother	317	7	0.5	218	2	G70654	hypothetical prote
245	7	0.5	171	2	E97386	hypothetical prote	318	7	0.5	219	2	S71472	endo-1,4-beta-xyla
246	7	0.5	171	2	AE2604	conserved hypother	319	7	0.5	219	2	T38809	probable lipote-p
247	7	0.5	174	2	E69038	heat shock protein	320	7	0.5	220	2	F64203	glycerol uptake fa
248	7	0.5	174	2	T46382	hypothetical prote	c 321	7	0.5	221	2	T36767	probable two-compo
249	7	0.5	175	2	T45369	ribosomal protein	c 322	7	0.5	221	2	AH2510	hypothetical prote
250	7	0.5	175	2	T28050	hypothetical prote	323	7	0.5	221	2	C34768	ORF2 protein - Orf
251	7	0.5	176	2	D87299	acetyltransferase,	c 324	7	0.5	222	2	G83125	conserved hypother
252	7	0.5	176	2	C55208	socA3 protein - My	325	7	0.5	223	2	I59173	glutamate decarbox
253	7	0.5	177	1	S74917	adenyl-yl-sulfate k	326	7	0.5	223	2	A36462	probable transmemb
254	7	0.5	177	2	S54778	NR-13 protein - qu	c 327	7	0.5	224	2	H97711	hypothetical prote
255	7	0.5	177	2	F72715	hypothetical prote	328	7	0.5	224	2	A26569	exotoxin A regulat
256	7	0.5	178	2	F82352	ATP synthase delta	c 329	7	0.5	226	2	T11485	H+-transporting tw
257	7	0.5	178	2	T00644	hypothetical prote	c 330	7	0.5	226	2	G83336	probable two-compo
258	7	0.5	178	2	E75340	conserved hypother	331	7	0.5	227	2	H86940	probable membrane
259	7	0.5	178	2	S31955	penicillin-binding	c 332	7	0.5	228	1	LCMS	prolactin precurs
260	7	0.5	178	2	AH2592	conserved hypother	333	7	0.5	228	2	JC6320	superoxide dismuta
261	7	0.5	179	2	G90003	ATP synthase delta	334	7	0.5	228	2	C75463	hypothetical prote
262	7	0.5	180	2	F97402	hypothetical prote	335	7	0.5	228	2	G86882	hypothetical prote
263	7	0.5	180	2	F70509	hypothetical prote	c 336	7	0.5	229	1	G64371	conserved hypother
264	7	0.5	180	2	T20378	hypothetical prote	c 337	7	0.5	229	2	AD2462	hypothetical prote
265	7	0.5	180	2	C88465	protein B0244.9 [1	c 338	7	0.5	230	2	T33622	hypothetical prote
266	7	0.5	180	2	B86796	hypothetical prote	c 339	7	0.5	231	2	A42471	uroporphyrin-III C
267	7	0.5	180	2	H82480	probable acetyltra	c 340	7	0.5	232	1	JH0597	transcription fact
268	7	0.5	180	2	B97375	hypothetical 25.1k	c 341	7	0.5	232	2	A83396	probable COA trans
269	7	0.5	181	2	T48558	hypothetical prote	c 342	7	0.5	232	2	T21526	hypothetical prote
270	7	0.5	182	2	D83530	cytochrome b561 PA	c 343	7	0.5	232	2	T31524	hypothetical prote
271	7	0.5	182	2	T16472	hypothetical prote	344	7	0.5	233	2	S60836	hypothetical prote
272	7	0.5	183	2	B29017	zein zcl - maize	c 345	7	0.5	234	2	S60885	ferric exochelin u
273	7	0.5	185	2	C86705	hypothetical prote	346	7	0.5	235	2	PQ0773	4-coumarate-CoA li
274	7	0.5	186	2	H69090	conserved hypother	347	7	0.5	235	2	S48924	hypothetical prote
275	7	0.5	186	2	G86492	polymorphic outer	348	7	0.5	235	2	D84602	hypothetical prote
276	7	0.5	187	2	S07271	p-aminobenzoate sy	c 349	7	0.5	235	2	A96575	hypothetical prote
277	7	0.5	187	2	T06581	probable deoxyribo	c 350	7	0.5	237	2	S55378	serine proteinase
278	7	0.5	188	2	S31952	penicillin-binding	351	7	0.5	237	2	F43692	T4 protein - rabbi
279	7	0.5	188	2	C87649	OmpA family protei	352	7	0.5	237	2	H72225	conserved hypother
280	7	0.5	189	2	D69886	hypothetical prote	c 353	7	0.5	237	2	T47755	hypothetical prote
281	7	0.5	189	2	AD0736	probable membrane	354	7	0.5	237	2	T25152	hypothetical prote
282	7	0.5	190	2	A47569	pl43 DNA helicase	c 355	7	0.5	238	2	F87605	hypothetical prote
283	7	0.5	190	2	T10740	carbonate dehydrat	c 356	7	0.5	239	2	G97553	sugar fermentation
284	7	0.5	190	2	A44455	prostaglandin-D sy	c 357	7	0.5	239	2	A12773	sugar fermentation
285	7	0.5	192	2	G83409	hypothetical prote	c 358	7	0.5	240	2	H95873	probable amino aci
286	7	0.5	193	2	H82028	hypothetical prote	c 359	7	0.5	240	2	AC1921	hypothetical prote
287	7	0.5	195	2	C75490	hypothetical prote	c 360	7	0.5	241	2	T47752	superoxide dismuta
288	7	0.5	197	2	E70642	probable ribosomal	c 361	7	0.5	241	2	B69090	hypothetical prote
289	7	0.5	197	2	A70033	carbonic anhydrase	c 362	7	0.5	241	2	F69150	conserved hypother
290	7	0.5	197	2	G81131	conserved hypother	c 363	7	0.5	241	2	F84233	hypothetical prote
291	7	0.5	197	2	A86430	F26G16.12 protein	364	7	0.5	241	2	A87409	conserved hypother
292	7	0.5	198	2	D97782	hypothetical prote	c 365	7	0.5	242	2	E88250	protein T21B10.5 l
293	7	0.5	198	2	S42134	light-harvesting c	c 366	7	0.5	243	2	AB1191	B. subtilis Yhf1 p
294	7	0.5	198	2	T33031	hypothetical prote	c 367	7	0.5	243	2	AB1549	conserved hypother
295	7	0.5	200	2	C89956	30S ribosomal prot	c 368	7	0.5	244	2	AE3507	amidotransferase h
296	7	0.5	200	2	AD3633	hypothetical prote	c 369	7	0.5	245	2	T43036	probable transaldo
297	7	0.5	201	2	J01094	hypothetical 20.2k	c 370	7	0.5	245	2	T51081	hypothetical prote
298	7	0.5	202	2	B87438	exonuclease [impor	c 371	7	0.5	246	2	S37341	chitinase (EC 3.2.
299	7	0.5	203	2	G97970	hypothetical prote	c 372	7	0.5	246	2	T01710	hypothetical prote
300	7	0.5	204	2	AB3385	leucyl/phenylalany	c 373	7	0.5	247	2	E55545	afab protein - Esc
301	7	0.5	204	2	E84749	hypothetical prote	c 374	7	0.5	247	2	S45063	hypothetical prote
302	7	0.5	205	2	C54759	bo-type ubiquinol	c 375	7	0.5	249	2	JH0629	cleavage signal-1
303	7	0.5	205	2	F64935	probable CDP-alcoh	c 376	7	0.5	249	2	S03173	exotoxin A regulat
304	7	0.5	208	2	D85785	probable cytochrom	377	7	0.5	250	1	GWBS	indole-3-glycerol-
305	7	0.5	208	2	E86221	hypothetical prote	378	7	0.5	250	2	E97385	cpaD protein (AF22
306	7	0.5	208	2	H90936	probable cytochrom	379	7	0.5	250	2	AE2603	components of type

c 380	7	0.5	250	2	H75355	hypothetical prote	453	7	0.5	280	2	S52479	hypothetical prote
c 381	7	0.5	250	2	E83820	hypothetical prote	454	7	0.5	280	2	AD1109	B. subtilis ComEC
c 382	7	0.5	251	1	B65188	Ubiquitinone/menai	455	7	0.5	280	2	AF1470	C-terminal part B.
c 383	7	0.5	251	2	C98224	hypothetical prote	456	7	0.5	281	2	PC6005	methionine-tRNA li
c 384	7	0.5	251	2	A86071	hypothetical prote	c 457	7	0.5	281	2	A72561	hypothetical prote
c 385	7	0.5	251	2	B64048	hypothetical prote	458	7	0.5	282	2	H84653	probable protein k
c 386	7	0.5	253	2	AG1122	B. subtilis transc	459	7	0.5	282	2	B96604	hypothetical prote
c 387	7	0.5	253	2	AI1482	B. subtilis transc	460	7	0.5	282	2	A56676	homeotic protein C
c 388	7	0.5	253	2	A83719	hypothetical prote	c 461	7	0.5	282	2	S38094	hypothetical prote
c 389	7	0.5	253	2	E64420	uroporphyrin-III C	462	7	0.5	283	2	S61664	vacuolar protein v
c 390	7	0.5	253	2	T01453	hypothetical prote	c 463	7	0.5	284	2	AC2346	malate dehydrogena
c 391	7	0.5	253	2	S76719	hypothetical prote	464	7	0.5	284	2	G72107	oligopeptide trans
c 392	7	0.5	254	2	H97948	hypothetical prote	c 465	7	0.5	284	2	I51172	transcription fact
c 393	7	0.5	254	2	F95081	hypothetical prote	466	7	0.5	284	2	D83959	hypothetical prote
c 394	7	0.5	255	2	AB3579	hypothetical cytos	c 467	7	0.5	284	2	S75817	hypothetical prote
c 395	7	0.5	255	2	F83409	hypothetical prote	468	7	0.5	284	2	A96604	hypothetical prote
c 396	7	0.5	256	2	F83223	conserved hypotnet	469	7	0.5	284	2	F81563	peptide ABC transp
c 397	7	0.5	256	2	S42932	probable transmem	470	7	0.5	284	2	A86516	oligopeptide trans
c 398	7	0.5	256	2	S42938	probable membrane-	c 471	7	0.5	285	2	S21562	hypothetical prote
c 399	7	0.5	257	2	A82691	tRNA pseudouridine	472	7	0.5	286	2	C97459	sulfate ABC transp
c 400	7	0.5	257	2	G42600	ABC-type transport	473	7	0.5	286	2	A82677	ABC transporter, m
c 401	7	0.5	257	2	AB3231	hypothetical prote	474	7	0.5	286	2	A82066	conserved hypotnet
c 402	7	0.5	257	2	C72651	hypothetical prote	c 475	7	0.5	286	2	F87075	membrane transport
c 403	7	0.5	258	2	T01481	carbonate dehydrat	476	7	0.5	286	2	A32349	chvD protein - Agr
c 404	7	0.5	258	2	D86371	hypothetical prote	c 477	7	0.5	287	1	B69865	probable phosphoes
c 405	7	0.5	259	1	TOH01	insulin-like growt	c 478	7	0.5	288	1	S41006	enoyl-CoA hydratase
c 406	7	0.5	259	2	T05756	hypothetical prote	479	7	0.5	288	2	G86223	hypothetical prote
c 407	7	0.5	259	2	D83557	transcription regu	480	7	0.5	288	2	T50082	azrl protein [impo
c 408	7	0.5	260	2	A87024	probable DNA-bind	c 481	7	0.5	288	2	C70860	hypothetical prote
c 409	7	0.5	260	2	T48846	creatininase (EC 3	482	7	0.5	288	2	S58219	ABA-inducible prot
c 410	7	0.5	262	2	A10053	probable metalloen	483	7	0.5	289	2	F64374	modification methy
c 411	7	0.5	262	2	T33597	hypothetical prote	484	7	0.5	289	2	D70809	probable pabc prot
c 412	7	0.5	263	1	S23009	insulin-like growt	485	7	0.5	290	2	G75470	conserved hypotnet
c 413	7	0.5	263	2	S72528	chitinase (EC 3.2.	486	7	0.5	291	2	S27721	hypothetical prote
c 414	7	0.5	263	2	S69184	chitinase (EC 3.2.	487	7	0.5	292	2	G88448	protein C4569.4 [i
c 415	7	0.5	263	2	D82441	conserved hypotnet	488	7	0.5	292	2	S40979	hypothetical prote
c 416	7	0.5	263	2	T01614	hypothetical prote	489	7	0.5	292	2	T26879	hypothetical prote
c 417	7	0.5	264	2	S26625	chitinase (EC 3.2.	490	7	0.5	292	2	S33513	gene Fif protein -
c 418	7	0.5	264	2	D97625	hypothetical prote	491	7	0.5	292	2	F97082	transcription regu
c 419	7	0.5	264	2	AF2848	conserved hypotnet	c 492	7	0.5	293	2	S47094	hypothetical prote
c 420	7	0.5	264	2	H89005	protein T24A6.18 [493	7	0.5	294	2	C36351	sarcotoxin II-3 -
c 421	7	0.5	265	2	H83358	probable permease	494	7	0.5	294	2	A27692	sarcotoxin IIA pre
c 422	7	0.5	265	2	A36351	sarcotoxin II-1 -	c 495	7	0.5	294	2	T29028	hypothetical prote
c 423	7	0.5	266	2	C81269	probable dimethyla	496	7	0.5	294	2	E88448	protein C4569.9 [i
c 424	7	0.5	268	2	T16544	hypothetical prote	c 497	7	0.5	294	2	T36070	hypothetical prote
c 425	7	0.5	268	2	T19829	hypothetical prote	498	7	0.5	294	2	T33931	hypothetical prote
c 426	7	0.5	268	2	A49303	homeotic protein C	c 499	7	0.5	296	2	T05110	hypothetical prote
c 427	7	0.5	269	1	A27067	calretinin - chick	500	7	0.5	299	2	AB0440	hypothetical prote
c 428	7	0.5	271	1	A60253	calretinin - human	c 501	7	0.5	301	1	A40811	myosin-light-chain
c 429	7	0.5	271	1	S25006	calretinin - rat	502	7	0.5	301	2	B90192	conserved hypotnet
c 430	7	0.5	271	2	H84337	spermidine/putresc	503	7	0.5	302	2	A32610	spermidine synthase
c 431	7	0.5	271	2	T48994	pirin-like protein	c 504	7	0.5	302	2	A72510	hypothetical prote
c 432	7	0.5	271	2	T01767	hypothetical prote	c 505	7	0.5	303	2	C84079	lipic acid synthet
c 433	7	0.5	271	2	S62485	probable mitochond	506	7	0.5	303	2	B83367	probable glycosyl
c 434	7	0.5	272	1	A36082	insulin-like growt	507	7	0.5	304	2	S04853	galactose 1-dehydr
c 435	7	0.5	272	2	I48600	insulin-like growt	508	7	0.5	305	2	H82080	UDP-3-O-3-hydroxym
c 436	7	0.5	272	2	T25044	hypothetical prote	509	7	0.5	305	2	A75211	asparaginase (EC 3
c 437	7	0.5	273	2	T22957	hypothetical prote	c 510	7	0.5	305	2	F64599	outer membrane pro
c 438	7	0.5	273	2	AF0686	hypothetical prote	c 511	7	0.5	306	1	W2WLEB	E2 protein - bovin
c 439	7	0.5	275	2	T03032	chitinase (EC 3.2.	c 512	7	0.5	306	2	C97347	oligopeptide ABC t
c 440	7	0.5	275	2	E69975	cyclodextrin metab	c 513	7	0.5	306	2	F97346	oligopeptide ABC t
c 441	7	0.5	275	2	A97226	uncharacterized co	c 514	7	0.5	306	2	C69749	conserved hypotnet
c 442	7	0.5	276	2	T08178	assemblin, striate	c 515	7	0.5	306	2	F70603	hypothetical prote
c 443	7	0.5	277	2	I37552	OX40 homolog - hum	c 516	7	0.5	306	2	G75597	hypothetical prote
c 444	7	0.5	277	2	G72357	sugar ABC transpor	c 517	7	0.5	308	2	H70767	hypothetical prote
c 445	7	0.5	278	2	AE0210	4-deoxy-L-threo-5-	518	7	0.5	309	2	T32376	hypothetical prote
c 446	7	0.5	278	2	D69959	conserved hypotnet	519	7	0.5	309	2	AI0314	probable sugar-bin
c 447	7	0.5	278	2	G87355	hypothetical prote	c 520	7	0.5	310	2	A86239	protein T10024.17
c 448	7	0.5	279	2	S76873	hypothetical prote	c 521	7	0.5	310	2	D95867	probable transcrip
c 449	7	0.5	280	2	F90845	formyltetrahydrofo	c 522	7	0.5	310	2	E64953	erfK protein precu
c 450	7	0.5	280	2	E85703	hypothetical prote	c 523	7	0.5	310	2	A99977	hypothetical prote
c 451	7	0.5	280	2	AF0649	formyltetrahydrofo	c 524	7	0.5	310	2	G85823	hypothetical prote
c 452	7	0.5	280	2	C36871	formyltetrahydrofo	c 525	7	0.5	310	2	C83088	hypothetical prote

c 526	7	0.5	311	2	JC2541	599	7	0.5	340	2	C97653	pseudouridine synt
527	7	0.5	311	2	H84746	c 600	7	0.5	340	2	AH1986	hypothetical prote
528	7	0.5	311	2	A83746	c 601	7	0.5	340	2	E95147	alcohol dehydrogen
529	7	0.5	312	2	A11871	c 602	7	0.5	341	2	A13425	ribose transport s
c 530	7	0.5	312	2	D87475	c 603	7	0.5	341	2	S43586	hypothetical prote
531	7	0.5	313	2	A65140	c 604	7	0.5	341	2	T04050	hypothetical prote
c 532	7	0.5	313	2	B96692	c 605	7	0.5	342	2	T29557	hypothetical prote
c 533	7	0.5	316	2	T14643	c 606	7	0.5	343	2	F83126	ferric enterobacti
534	7	0.5	316	2	T19396	c 607	7	0.5	344	2	AB3236	hypothetical prote
535	7	0.5	316	2	I38754	c 608	7	0.5	345	2	C45456	NADH2 dehydrogenas
c 536	7	0.5	317	2	S40535	c 609	7	0.5	345	2	S23368	NADH2 dehydrogenas
c 537	7	0.5	317	2	H90629	c 610	7	0.5	345	2	C38330	cation efflux syst
c 538	7	0.5	317	2	H85480	c 611	7	0.5	345	2	D84089	hypothetical prote
c 539	7	0.5	317	2	AH0502	c 612	7	0.5	345	2	AE2820	two component sens
c 540	7	0.5	317	2	AD0057	c 613	7	0.5	346	2	B69950	conserved hypothet
541	7	0.5	317	2	T39869	c 614	7	0.5	346	2	T09927	cytidine deaminase
542	7	0.5	317	2	G86761	c 615	7	0.5	346	2	A84398	hypothetical prote
c 543	7	0.5	317	2	S72851	c 616	7	0.5	346	2	C98015	conserved hypothet
544	7	0.5	317	2	T34228	c 617	7	0.5	346	2	ED7598	fsrR protein (AF19
c 545	7	0.5	319	1	LURT5	c 618	7	0.5	347	2	AD2711	aldo/keto reductas
546	7	0.5	319	2	T46145	c 619	7	0.5	348	2	T14040	NADH2 dehydrogenas
547	7	0.5	319	2	T36845	c 620	7	0.5	348	2	G02297	gene N33 protein -
c 548	7	0.5	319	2	AE2823	c 621	7	0.5	349	2	I39535	probable L-lactate
c 549	7	0.5	320	2	T42062	c 622	7	0.5	349	2	G84852	hypothetical prote
c 550	7	0.5	320	2	S38670	c 623	7	0.5	349	2	A53340	interferon regulat
551	7	0.5	321	2	G86010	c 624	7	0.5	351	2	A13355	UDP-3-O-[3-hydroxy
c 552	7	0.5	322	1	A53307	c 625	7	0.5	351	2	T51513	hypothetical prote
c 553	7	0.5	322	2	T36577	c 626	7	0.5	352	2	T38311	protein kinase - f
c 554	7	0.5	322	2	T40834	c 627	7	0.5	352	2	G00048	fusin (LESTRA) - c
555	7	0.5	323	2	JQ1552	c 628	7	0.5	352	2	A45747	neuropeptide Y/pep
556	7	0.5	324	1	D42951	c 629	7	0.5	352	2	T35031	probable transcrip
557	7	0.5	325	2	B90805	c 630	7	0.5	352	2	F83313	conserved hypothet
558	7	0.5	325	2	G85664	c 631	7	0.5	353	2	A10220	probable sugar tra
559	7	0.5	325	2	F64845	c 632	7	0.5	353	2	B72488	hypothetical prote
560	7	0.5	325	2	A12876	c 633	7	0.5	354	1	GNV5R	genome polyprotein
561	7	0.5	325	2	T39520	c 634	7	0.5	355	2	A42347	opsin, green-sensi
c 562	7	0.5	325	2	F90508	c 635	7	0.5	355	2	F90391	conserved hypothet
563	7	0.5	326	2	E64887	c 636	7	0.5	357	2	AB3516	sensory transducti
564	7	0.5	326	2	H82455	c 637	7	0.5	357	1	A54674	L-iditol 2-dehydro
c 565	7	0.5	326	2	T21924	c 638	7	0.5	360	1	A25732	inhibin alpha chai
c 566	7	0.5	326	2	G87698	c 639	7	0.5	360	2	S69063	probable membrane
c 567	7	0.5	329	2	S28305	c 640	7	0.5	360	2	T27324	hypothetical prote
568	7	0.5	329	2	T00873	c 641	7	0.5	361	2	T14800	hypothetical prote
569	7	0.5	329	2	AF0019	c 642	7	0.5	361	2	A75195	d-aminopeptidase
570	7	0.5	331	2	G91164	c 643	7	0.5	361	2	D71227	probable D-aminope
571	7	0.5	331	2	AE0481	c 644	7	0.5	362	2	AE1397	conserved hypothet
572	7	0.5	331	2	AB0995	c 645	7	0.5	362	2	AH1772	conserved hypothet
c 573	7	0.5	332	2	S63660	c 646	7	0.5	363	2	T47588	hypothetical prote
c 574	7	0.5	332	2	AF3265	c 647	7	0.5	364	1	WFFGA	hypothetical prote
c 575	7	0.5	332	2	S77370	c 648	7	0.5	364	2	T03892	inhibin alpha chai
c 576	7	0.5	333	2	F90225	c 649	7	0.5	364	2	T25124	hypothetical prote
c 577	7	0.5	333	2	S78136	c 650	7	0.5	365	2	E83952	1-deoxy-d-xylulose
c 578	7	0.5	333	2	F70672	c 651	7	0.5	365	2	E83600	hypothetical prote
579	7	0.5	333	2	T05187	c 652	7	0.5	366	1	A24248	inhibin alpha chai
c 580	7	0.5	333	2	E86444	c 653	7	0.5	366	1	JC1106	inhibin alpha chai
c 581	7	0.5	333	2	G86237	c 654	7	0.5	366	1	A40056	inhibin alpha chai
582	7	0.5	335	2	A72417	c 655	7	0.5	366	1	A40056	inhibin alpha chai
583	7	0.5	335	2	T44728	c 656	7	0.5	366	2	A96692	hypothetical prote
c 584	7	0.5	335	2	C65134	c 657	7	0.5	367	2	E83676	pyruvate dehydroge
c 585	7	0.5	335	2	A34290	c 658	7	0.5	367	2	S31651	MHC class I histoc
586	7	0.5	336	2	T09186	c 659	7	0.5	367	2	S25185	hypothetical prote
587	7	0.5	336	2	S61299	c 660	7	0.5	367	2	T00580	probable facyl-car
588	7	0.5	336	2	B81869	c 661	7	0.5	368	2	A82249	response regulator
589	7	0.5	336	2	C81073	c 662	7	0.5	369	2	AG3502	phenylalanine-tRNA
c 590	7	0.5	336	2	S55863	c 663	7	0.5	369	2	G71831	co-chaperone with
591	7	0.5	337	2	T24387	c 664	7	0.5	369	2	D64686	co-chaperone and h
592	7	0.5	337	2	T23794	c 665	7	0.5	369	2	T18663	hypothetical prote
593	7	0.5	338	1	IEEC5D	c 666	7	0.5	369	2	D97601	hypothetical prote
594	7	0.5	338	2	A64964	c 667	7	0.5	370	2	B97493	hypothetical prote
595	7	0.5	338	2	C70553	c 668	7	0.5	371	2	A23431	NADH2 dehydrogenas
596	7	0.5	338	2	F69035	c 669	7	0.5	371	2	E97396	glutathione synthet
597	7	0.5	338	2	A89397	c 670	7	0.5	372	2	A95338	hypothetical prote
598	7	0.5	339	2	S64381	c 671	7	0.5	373	2	I69008	MHC class I RT1.E

672	7	0.5	373	2	E95102	hypothetical prote	c 745	7	0.5	409	2	D96912	o-acetylthomoserine
c 673	7	0.5	373	2	JC2426	transcription acti	c 746	7	0.5	409	2	T46741	arginine delaminase
674	7	0.5	374	1	A43407	cysteine synthase	c 747	7	0.5	409	2	F71828	DNA transfer prote
c 675	7	0.5	375	2	S76663	hypothetical prote	c 748	7	0.5	409	2	H82961	conserved hypothet
c 676	7	0.5	375	2	T41885	ODV-E56 orf148 - B	c 749	7	0.5	409	2	T50311	oxalisp2 protein [i
677	7	0.5	376	2	A55874	DNA-directed DNA p	c 750	7	0.5	409	2	T43703	respiratory protei
c 678	7	0.5	376	2	B96714	hypothetical prote	c 751	7	0.5	410	2	E86879	arginine delaminase
c 679	7	0.5	376	2	T42673	hypothetical prote	c 752	7	0.5	410	2	F82314	conserved hypothet
680	7	0.5	376	2	C82847	conserved hypothet	c 753	7	0.5	411	1	F71178	translation initia
681	7	0.5	377	2	B97376	cytochrome ba(3) (c 754	7	0.5	411	1	E75163	translation initia
682	7	0.5	377	2	A45851	MHC class I histoc	c 755	7	0.5	411	2	H70908	hypothetical prote
c 683	7	0.5	377	2	T04585	hypothetical prote	c 756	7	0.5	413	2	T23098	hypothetical prote
c 684	7	0.5	378	2	C64091	adenine glycosylas	c 757	7	0.5	413	2	G95305	conserved hypothet
685	7	0.5	379	2	E75460	conserved hypothet	c 758	7	0.5	414	2	A40350	transcription repr
c 686	7	0.5	380	2	D64129	probable 8-amino-7	c 759	7	0.5	414	2	A48273	delta/VY1/NF-EJ/UC
c 687	7	0.5	380	2	S61598	probable membrane	c 760	7	0.5	417	2	JC4698	divalent cation re
688	7	0.5	380	2	T04534	hypothetical prote	c 761	7	0.5	418	2	T51814	polypyrimidine tra
689	7	0.5	380	2	S70964	pkn5 protein - Myx	c 762	7	0.5	419	2	A25438	keratin, type I cy
c 690	7	0.5	381	1	B69030	conserved hypothet	c 763	7	0.5	419	2	F83673	PTS system, galact
c 691	7	0.5	382	1	DEKVG	glycerate dehydrog	c 764	7	0.5	419	2	A40728	microphthalma-ass
c 692	7	0.5	382	2	B88561	protein F58A4.7b [c 765	7	0.5	420	2	JC4716	zinc finger DNA-bi
c 693	7	0.5	383	2	H75263	probable zinc prot	c 766	7	0.5	420	2	AH3015	hemolysin [importe
c 694	7	0.5	384	2	A47249	brain-specific som	c 767	7	0.5	422	1	W2WLB2	E2 protein - Bovin
c 695	7	0.5	385	2	S54992	reverse transcript	c 768	7	0.5	422	2	G88481	protein Cl6A3.9 [i
696	7	0.5	385	2	F15K9.6 [i	protein F15K9.6 [i	c 769	7	0.5	423	2	PQ0772	4-coumarate-CoA li
697	7	0.5	385	2	T18180	proline-rich prote	c 770	7	0.5	425	2	H90415	hypothetical prote
c 698	7	0.5	386	2	G89801	hypothetical prote	c 771	7	0.5	425	2	T50355	hypothetical prote
c 699	7	0.5	387	2	C83398	hypothetical prote	c 772	7	0.5	425	2	G98268	hemolysin (AF21204
700	7	0.5	387	2	A82849	hypothetical prote	c 773	7	0.5	425	2	S55147	hypothetical prote
c 701	7	0.5	388	2	JN0605	somatostatin recep	c 774	7	0.5	427	2	G81260	UDP-N-acetylmuramo
c 702	7	0.5	388	2	S75920	hypothetical prote	c 775	7	0.5	428	2	A43741	terminus protein -
703	7	0.5	389	2	F71865	cyclopropane-fatty	c 776	7	0.5	429	2	F91175	arsenical pump mem
704	7	0.5	389	2	H64571	cyclopropane-fatty	c 777	7	0.5	429	2	F86021	arsenical pump mem
705	7	0.5	389	2	B61380	hypothetical prote	c 778	7	0.5	429	2	D69635	H+/Na+-glutamate s
c 706	7	0.5	389	2	A12230	mannose-1-phosphat	c 779	7	0.5	430	2	E64059	probable isochoris
707	7	0.5	389	2	AG2460	hypothetical prote	c 780	7	0.5	430	2	T05980	hypothetical prote
708	7	0.5	390	2	A64549	conserved hypothet	c 781	7	0.5	431	1	WJHU2G	homeotic protein H
709	7	0.5	390	2	E71958	hypothetical prote	c 782	7	0.5	431	2	A72294	adenylosuccinate l
710	7	0.5	392	1	XNHUSP	serine-pyruvate tr	c 783	7	0.5	431	2	S09824	hypothetical prote
711	7	0.5	392	2	S24155	alanine-glyoxylate	c 784	7	0.5	432	2	AE1170	hypothetical prote
c 712	7	0.5	393	2	A71201	hypothetical prote	c 785	7	0.5	432	2	AG1527	hypothetical prote
c 713	7	0.5	393	2	B86189	protein T25N20.9 [c 786	7	0.5	432	2	AB2222	twitching motility
c 714	7	0.5	395	2	T02185	probale translatio	c 787	7	0.5	433	2	S20963	homeotic protein H
715	7	0.5	396	2	T18854	hypothetical prote	c 788	7	0.5	433	2	S69999	sterigmatocystin s
c 716	7	0.5	396	2	S58161	probable translati	c 789	7	0.5	433	2	B87153	phoH-like protein
c 717	7	0.5	396	2	T04224	hypothetical prote	c 790	7	0.5	433	2	E70896	probable phoH2 pro
718	7	0.5	398	2	T04182	hypothetical prote	c 791	7	0.5	435	2	T00126	hypothetical prote
c 719	7	0.5	398	2	T02484	hypothetical prote	c 792	7	0.5	436	2	S47723	arsenite efflux pu
720	7	0.5	398	2	E87403	hypothetical prote	c 793	7	0.5	436	2	B48957	cysteine aminopept
721	7	0.5	398	2	C97679	bcr protein (AE006	c 794	7	0.5	436	2	F86859	aminopeptidase C [
c 722	7	0.5	398	2	A12903	MFS permease [drug	c 795	7	0.5	436	2	JC5021	platelet-activatin
c 723	7	0.5	399	1	B49836	transcription fact	c 796	7	0.5	436	2	S64744	hypothetical prote
724	7	0.5	399	2	C83611	hypothetical prote	c 797	7	0.5	436	2	F71288	probable sugar ABC
c 725	7	0.5	400	1	S22539	transcription fact	c 798	7	0.5	436	2	A12340	hypothetical prote
c 726	7	0.5	400	2	A60198	lipoprotein releas	c 799	7	0.5	437	2	A12340	hypothetical prote
727	7	0.5	401	2	F87196	probable membrane-	c 800	7	0.5	437	2	S59151	NADH2 dehydrogenas
c 728	7	0.5	401	2	T35420	probable secreted	c 801	7	0.5	437	2	T39578	nccC protein - Alc
729	7	0.5	401	2	T01001	hypothetical prote	c 802	7	0.5	437	2	T14192	extensin homolog T
730	7	0.5	401	2	AE3651	amidase (EC 3.5.1.	c 803	7	0.5	439	1	DEPSHA	homoserine dehydro
731	7	0.5	402	2	G83128	probable MFS trans	c 804	7	0.5	440	1	B70326	conserved hypothet
732	7	0.5	402	2	S32142	hypothetical prote	c 805	7	0.5	440	2	T35925	diaminopimelate de
c 733	7	0.5	403	2	AF3432	site-specific DNA-	c 806	7	0.5	441	2	E75357	probable periplasm
734	7	0.5	403	2	T47621	bZIP transcription	c 807	7	0.5	442	2	G71930	probable transport
735	7	0.5	403	2	C87315	hypothetical prote	c 808	7	0.5	442	2	B64582	sodium- and chlori
736	7	0.5	403	2	D87597	hypothetical prote	c 809	7	0.5	444	1	F64138	argininosuccinate
737	7	0.5	404	2	H87635	Rieske 2Fe-2S fami	c 810	7	0.5	444	2	B95033	aminopeptidase C [
738	7	0.5	404	2	H64160	hypothetical prote	c 811	7	0.5	444	2	B97904	aminopeptidase C (
739	7	0.5	404	2	H64175	hypothetical prote	c 812	7	0.5	444	2	T42674	hypothetical prote
c 740	7	0.5	404	2	T50335	mating pheromone r	c 813	7	0.5	445	2	JC2525	UDP-glucose dehydr
c 741	7	0.5	405	2	E69977	conserved hypothet	c 814	7	0.5	445	2	T11097	NADH2 dehydrogenas
c 742	7	0.5	406	2	A39339	protein C inhibito	c 815	7	0.5	446	2	A82337	CDP-diacetylglcerol
c 743	7	0.5	406	2	T36632	probable oxidoredu	c 816	7	0.5	446	2	T25793	hypothetical prote
c 744	7	0.5	408	2	T29050	delta-1-pyrroline-	c 817	7	0.5	446	2	A38244	citrate transport

c 818	7	0.5	446	2	A42661	citrate carrier pr	c 891	7	0.5	478	1	I47154	transcription fact
c 819	7	0.5	446	2	B42661	citrate carrier pr	c 892	7	0.5	478	2	AF1758	chitinase and chit
c 820	7	0.5	446	2	AD0509	citrate-sodium sym	c 893	7	0.5	479	1	A31753	transcription fact
c 821	7	0.5	447	2	F82025	argininosuccinate	c 894	7	0.5	479	1	S22542	transcription fact
c 822	7	0.5	447	2	B81003	argininosuccinate	c 895	7	0.5	480	2	AE1121	beta-glucosidase h
c 823	7	0.5	448	2	B72262	glucose-6-phosphat	c 896	7	0.5	480	2	T46047	hypothetical prote
c 824	7	0.5	448	2	A83091	tryptophanyl-tRNA	c 897	7	0.5	481	2	E86356	hypothetical prote
c 825	7	0.5	449	1	A41386	clusterin precursor	c 898	7	0.5	481	2	A33712	metalloproteinase
c 826	7	0.5	449	2	S67819	GumC protein - Xan	c 899	7	0.5	482	2	S31478	alpha-amylase (EC
c 827	7	0.5	449	2	A84900	hypothetical prote	c 900	7	0.5	482	2	S77660	ribosomal protein
c 828	7	0.5	449	2	T50555	delta-8 spinninglip	c 901	7	0.5	482	2	AF1120	conserved hypoteth
c 829	7	0.5	450	2	A54429	paired box transcr	c 902	7	0.5	485	1	S22543	transcription fact
c 830	7	0.5	450	2	T08701	hypothetical prote	c 903	7	0.5	485	1	AD0041	ramnulokinase (EC
c 831	7	0.5	452	2	F83587	inner membrane pro	c 904	7	0.5	485	2	T13694	glucuronosyltransf
c 832	7	0.5	452	2	S77436	sigma factor sibg	c 905	7	0.5	487	1	S52261	NADH2 dehydrogenas
c 833	7	0.5	452	2	T40769	hypothetical prote	c 906	7	0.5	487	1	QOECRS	vgJE protein - Esc
c 834	7	0.5	452	2	T34542	hypothetical prote	c 907	7	0.5	487	2	B91122	hypothetical prote
c 835	7	0.5	452	2	AE1710	hypothetical prote	c 908	7	0.5	487	2	A85967	hypothetical prote
c 836	7	0.5	452	2	AG1339	hypothetical prote	c 909	7	0.5	488	2	I56507	histamine H1 recep
c 837	7	0.5	453	2	F88445	protein C26E6.3 [i	c 910	7	0.5	488	2	A81006	probable malate de
c 838	7	0.5	454	2	H70158	conserved hypoteth	c 911	7	0.5	488	2	B82029	probable malate de
c 839	7	0.5	455	2	AE0191	argininosuccinate	c 912	7	0.5	488	2	B95332	hypothetical prote
c 840	7	0.5	455	2	F71477	probable D-Ala/Gly	c 913	7	0.5	489	2	D87551	glutamyl-tRNA(Gln)
c 841	7	0.5	455	2	F83460	probable amino aci	c 914	7	0.5	490	2	S71776	calcium-dependent
c 842	7	0.5	455	2	JC1224	nucleobindin precu	c 915	7	0.5	490	2	PQ0164	sucrose phosphoryl
c 843	7	0.5	456	2	S30922	ferredoxin-nitrite	c 916	7	0.5	490	2	T05444	hypothetical prote
c 844	7	0.5	457	2	A56925	paired box transcr	c 917	7	0.5	491	1	I5ECKR	ketol-acid reducto
c 845	7	0.5	457	2	E88456	protein W03A5.3 [i	c 918	7	0.5	491	1	EDBEM5	immediate-early pr
c 846	7	0.5	458	2	G83690	hypothetical prote	c 919	7	0.5	491	2	D91217	ketol-acid reducto
c 847	7	0.5	459	2	T11463	NADH2 dehydrogenas	c 920	7	0.5	491	2	E86063	ketol-acid reducto
c 848	7	0.5	459	2	S52250	paired box transcr	c 921	7	0.5	491	2	AF0923	ketol-acid reducto
c 849	7	0.5	459	2	T04808	hypothetical prote	c 922	7	0.5	491	2	AH3416	AMP nucleosidase (
c 850	7	0.5	459	2	I55472	Calcium binding pr	c 923	7	0.5	493	1	S17663	NADH2 dehydrogenas
c 851	7	0.5	460	2	C27311	NADH2 dehydrogenas	c 924	7	0.5	493	2	AG2738	glutamyl-tRNA amid
c 852	7	0.5	460	2	T09956	NADH2 dehydrogenas	c 925	7	0.5	494	2	AI1480	conserved hypoteth
c 853	7	0.5	460	2	T11773	NADH2 dehydrogenas	c 926	7	0.5	495	1	P1WLB	L1 protein - bovin
c 854	7	0.5	460	2	T11543	NADH2 dehydrogenas	c 927	7	0.5	496	2	T22405	protein-tyrosine k
c 855	7	0.5	460	2	T11309	NADH2 dehydrogenas	c 928	7	0.5	496	2	E97519	glutamyl-tRNA(gln)
c 856	7	0.5	460	2	S35470	NADH2 dehydrogenas	c 929	7	0.5	496	2	T41114	uroporphyrin methy
c 857	7	0.5	460	2	A58843	NADH2 dehydrogenas	c 930	7	0.5	498	2	B90456	hypothetical prote
c 858	7	0.5	460	2	B58893	NADH2 dehydrogenas	c 931	7	0.5	499	2	S70113	hypothetical prote
c 859	7	0.5	460	2	T09866	NADH2 dehydrogenas	c 932	7	0.5	501	1	P1WLB2	L1 protein - bovin
c 860	7	0.5	461	2	T19111	hypothetical prote	c 933	7	0.5	502	2	F64543	conserved hypoteth
c 861	7	0.5	461	2	T11829	NADH2 dehydrogenas	c 934	7	0.5	503	2	E83642	choline sulfatase
c 862	7	0.5	461	2	S57713	probable mannosyl	c 935	7	0.5	504	2	AE3526	5-carboxymethyl-2-
c 863	7	0.5	462	2	T11136	NADH2 dehydrogenas	c 936	7	0.5	504	2	T21377	hypothetical prote
c 864	7	0.5	463	2	H97952	sodium-dependent t	c 937	7	0.5	504	2	S17248	regulatory protein
c 865	7	0.5	463	2	T25193	hypothetical prote	c 938	7	0.5	505	2	T02898	hypothetical prote
c 866	7	0.5	464	2	AF2180	diaminopimelate de	c 939	7	0.5	506	2	T19787	hypothetical prote
c 867	7	0.5	465	2	E71409	sulfate adenyllylr	c 940	7	0.5	506	2	D81703	conserved hypoteth
c 868	7	0.5	465	2	D82262	probable capsular	c 941	7	0.5	507	2	S44916	2K688.6 protein -
c 869	7	0.5	465	2	T26852	hypothetical prote	c 942	7	0.5	508	2	S73540	phosphoglycerate m
c 870	7	0.5	465	2	T27885	sperm membrane pro	c 943	7	0.5	509	2	B90573	hypothetical prote
c 871	7	0.5	466	2	T30040	hypothetical prote	c 944	7	0.5	509	2	H87287	conserved hypoteth
c 872	7	0.5	467	2	H69109	heat shock protein	c 945	7	0.5	511	1	E89775	2',3'-cyclic-nucle
c 873	7	0.5	468	2	H87044	probable solute-bi	c 946	7	0.5	511	2	D64905	probable sugar tra
c 874	7	0.5	468	2	B70952	probable lpgy prot	c 947	7	0.5	511	2	A85724	hypothetical prote
c 875	7	0.5	469	2	H7467	glycosyl hydrolase	c 948	7	0.5	511	2	AC0941	probable ABC trans
c 876	7	0.5	469	2	T45201	hypothetical prote	c 949	7	0.5	511	2	H90893	hypothetical prote
c 877	7	0.5	470	2	H82054	conserved hypoteth	c 950	7	0.5	511	2	T37681	kinesin-like prote
c 878	7	0.5	471	2	C71439	hypothetical prote	c 951	7	0.5	512	2	B87804	protein K04F10.2 [
c 879	7	0.5	471	2	A84741	probable myrosinas	c 952	7	0.5	512	2	PT0764	endoooligopeptidase
c 880	7	0.5	471	2	AE3384	zinc metalloprotei	c 953	7	0.5	514	1	JT0565	IMP dehydrogenase
c 881	7	0.5	472	1	A49836	transcription fact	c 954	7	0.5	514	1	MMECMF	maltose transport
c 882	7	0.5	472	2	G81325	Glu-tRNAgin amidot	c 955	7	0.5	514	2	S05332	inner membrane pro
c 883	7	0.5	473	2	D83611	probable transcrip	c 956	7	0.5	514	2	S20604	transport system p
c 884	7	0.5	474	1	VYHUD	vitamin D-binding	c 957	7	0.5	514	2	H91255	part of maltose pe
c 885	7	0.5	474	2	F83550	serine proteinase	c 958	7	0.5	514	2	D86096	maltose transport
c 886	7	0.5	475	2	T11586	hypothetical prote	c 959	7	0.5	514	2	AC1014	maltose transport
c 887	7	0.5	475	2	T28966	hypothetical prote	c 960	7	0.5	515	2	G83004	phosphoglycerate m
c 888	7	0.5	476	2	AC2465	6-phosphogluconate	c 961	7	0.5	515	2	S56784	hypothetical prote
c 889	7	0.5	476	2	T06792	adenylosuccinate s	c 962	7	0.5	516	2	T44002	conserved herpesvi
c 890	7	0.5	476	2	E83006	two-component resp	c 963	7	0.5	517	2	T29852	hypothetical prote

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKC1 (1-393)

QY 1680 CTCCTCCACCCCTCAGAGCTCCATC 1706

DB 251 LeuProThrProLeuArgAlaSerIle 259

RESULT 4

HVRKC2

Ig mu chain C region (clone 12022) - horn shark (fragment)

C:Species: Heterodontus francisci (horn shark)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

C:Accession: S00980

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

EMBO J. 7, 1979-1988, 1988

A:Title: Complete structure and organization of immunoglobulin heavy chain constant region

A:Reference number: S00980; MUID:88328985; PMID:3138109

A:Accession: S00980

A:Molecule type: mRNA

A:Residues: 1-438 <KOK>

A:Cross-references: EMBL:X07784; NID:g63963; PIDN:CAA30617.1; PID:g63964

A:Note: the sequence was determined from the differentiated gene

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)

chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin

F:1-438/Domain: C region <CRE>

F:20-87/Domain: immunoglobulin homology <IM1>

F:123-190/Domain: immunoglobulin homology <IM2>

F:228-291/Domain: immunoglobulin homology <IM3>

F:330-400/Domain: immunoglobulin homology <IM4>

F:166,200,245,275,374,411,415,425/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.: 6.81 Length: 438

Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.68% Indels: 0

DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKC2 (1-438)

QY 1680 CTCCTCCACCCCTCAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 5

HVRKCS

Ig mu chain C region, secreted (clone 3050) - horn shark

C:Species: Heterodontus francisci (horn shark)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

C:Accession: S01853

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

EMBO J. 7, 1979-1988, 1988

A:Title: Complete structure and organization of immunoglobulin heavy chain constant region

A:Reference number: S00980; MUID:88328985; PMID:3138109

A:Accession: S01853

A:Molecule type: DNA

A:Residues: 1-438 <KOK>

A:Cross-references: EMBL:X07781

A:Note: the sequence was determined from the germline gene

C:Genetics:

A:Introns: 100/1; 206/1; 309/1

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)

chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin

F:1-438/Domain: C region <CRE>

F:20-87/Domain: immunoglobulin homology <IM1>

F:123-190/Domain: immunoglobulin homology <IM2>
F:228-291/Domain: immunoglobulin homology <IM3>
F:330-400/Domain: immunoglobulin homology <IM4>
F:164,200,245,275,374,411,415,425/Binding site: carbohydrate (Asn) (covalent) #status

Alignment Scores:

Pred. No.: 6.81 Length: 438

Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.68% Indels: 0

DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKCS (1-438)

QY 1680 CTCCTCCACCCCTCAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 6

HVRKC0

Ig mu chain C region, membrane-bound (clone 3050) - horn shark

C:Species: Heterodontus francisci (horn shark)

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Aug-1996

C:Accession: S01854; C32716; A46530

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

EMBO J. 7, 1979-1988, 1988

A:Title: Complete structure and organization of immunoglobulin heavy chain constant region

A:Reference number: S00980; MUID:88328985; PMID:3138109

A:Accession: S01854

A:Molecule type: DNA

A:Residues: 1-461 <KOK>

A:Cross-references: EMBL:X07781

A:Note: the sequence was determined from the germline gene

R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloott, M.J.; Litman, G.W.

Proc. Natl. Acad. Sci. U.S.A. 84, 5868-5872, 1987

A:Title: Extensive families of constant region genes in a phylogenetically primitive

A:Reference number: A32716; MUID:87289703; PMID:3475706

A:Accession: C32716

A:Molecule type: DNA

A:Residues: 1-99 <KO2>

A:Cross-references: GB:M17186

C:Genetics:

A:Introns: 100/1; 206/1; 309/1; 419/1; 459/3

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)

chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin

F:20-87/Domain: immunoglobulin homology <IMM1>

F:123-190/Domain: immunoglobulin homology <IMM2>

F:228-291/Domain: immunoglobulin homology <IMM3>

F:330-400/Domain: immunoglobulin homology <IMM4>

F:438-458/Domain: transmembrane #status predicted <TMW>

F:27-85,130-188,235-289,337-398/Disulfide bonds: #status predicted

F:164,200,245,275,374,411,415,437/Binding site: carbohydrate (Asn) (covalent) #status

Alignment Scores:

Pred. No.: 6.78 Length: 461

Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.68% Indels: 0

DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x HVRKC0 (1-461)

QY 1680 CTCCTCCACCCCTCAGAGCTCCATC 1706

DB 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 7

C84091

hypothetical protein BH3531 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84091
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C84091
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1490 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:gl0176109; PIDN:BA07250.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3531

Alignment Scores:
Pred. No.: 6.75 Length: 490
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.70% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x C84091 (1-490)

QY 202 TCAGAAAGCTGAAGATCTGGAAGGT 176

Db 398 SerGluArgLeuLysAspLeuGluGly 406

RESULT 8

JC5669

Ca2+/calmodulin-dependent protein kinase (EC 2.7.1.-) beta chain - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 20-Nov-1997 #sequence_revision 20-Nov-1997 #text_change 24-Sep-1999

C:Accession: JC5669; PC4493

R:Kitani, T.; Okuno, S.; Fujisawa, H.

J. Biochem. 122, 243-250, 1997

A:Title: Molecular cloning of Ca2+/calmodulin-dependent protein kinase kinase beta.

A:Reference number: JC5669; MUID:97420710; PMID:9276695

A:Accession: JC5669

A:Molecule type: mRNA

A:Residues: 1-587 <KIT>

A:Cross-references: GB:AB018081; NID:g3702720; PIDN:BAA33524.1; PID:d1034490; PID:g37027

A:Experimental source: cerebellar

A:Accession: PC4493

A:Molecule type: protein

A:Residues: 425-501 <KIT>

A:Experimental source: cerebellar

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C:Keywords: ATP; phosphotransferase

F:162-445/Domain: protein kinase homology <KIN>

F:170-194/Domain: ATP-binding #status predicted <ATP>

F:480-493/Domain: calmodulin-binding #status predicted <CAB>

Alignment Scores:
Pred. No.: 6.65 Length: 587
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x JC5669 (1-587)

QY 2180 TCCCTCCAGCAACACACCCAGCAA 2206

Db 571 SerProProArgThrProProGlnGln 579

RESULT 9

S67283

hypothetical protein YOR371c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O6679

C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
C:Accession: S67283
R:Dalius, H.; Hebling, U.; Hofmann, B.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67261
A:Accession: S67283
A:Molecule type: DNA
A:Residues: 1-897
A:Cross-references: EMBL:Z75279; NID:gl420801; PID:gl420802; GSPDB:GN00015; MIPS:YOR3
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:GPE1; MIPS:YOR371c
A:Cross-references: SGD:S0005898
A:Map position: 15R

Alignment Scores:
Pred. No.: 6.44 Length: 897
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.70% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x S67283 (1-897)

QY 2795 ATTTTCATCATCTGATTCATCAACAGGA 2769

Db 93 IleSerSerAspSerSerThrGly 101

RESULT 10

S45867

TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1

N:Alternate names: hypothetical protein YBR0207

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 22-Oct-1999

C:Accession: S45867

R:Entian, K.D.; Koetter, P.; Rose, M.; Li, Z.; Thermann, R.; Brendel, M.; Baur, A.; B

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45862

A:Accession: S45867

A:Molecule type: DNA

A:Residues: 1-1756 <ENT>

A:Cross-references: EMBL:Z35881; NID:g536207; PIDN:CAA84952.1; PID:g536210

A:Experimental source: strain S288C

C:Genetics:

A:Map position: 2R

A:Mobile element: retrotransposon Ty1

C:Superfamily: TyB protein

Alignment Scores:
Pred. No.: 6.12 Length: 1756
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x S45867 (1-1756)

QY 2779 GAATCAGATGATGAATTCACATGAT 2805

Db 1460 GluSerAspGluIleGlnHisAsp 1468

RESULT 11

F83374

hypothetical protein PA2174 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83374

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83374
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-102 <STO>
A:Cross-references: GB:AE004644; GB:AE004091; NID:g9948185; PIDN:AAG05562.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2174

Alignment Scores:
Pred. No.: 86 Length: 102
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x F83374 (1-102)
QY 586 AGCTCCAGTTTCAGATCTGCA 609
Db 70 SerSerPheSerAspLeuSer 77
|||||

RESULT 12
T44493
Conserved hypothetical protein YVAP [imported] - *Bacillus halodurans*
C:Species: *Bacillus halodurans*
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Jun-2001
C:Accession: T44493; G84062
R:Takami, H.; Takaki, Y.; Nakasone, K.; Sakiyama, T.; Maeno, G.; Sasaki, R.; Hiram, C.;
Extremophiles 3, 227-233, 1999
A:Title: Genetic analysis of the chromosome of alkaliphilic *Bacillus halodurans* C-125.
A:Reference number: 222781; MUID:99411980; PMID:10484179
A:Accession: T44493
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-107 <TAK>
A:Cross-references: EMBL:AB024562; NID:95822804; PIDN:BA83951.1; PID:95822806
A:Experimental source: strain C-125
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G84062
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA07022.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3303
A:Note: yvAP
C:Superfamily: conserved hypothetical protein MTH1285

Alignment Scores:
Pred. No.: 85.6 Length: 107
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x T44493 (1-107)
QY 202 TCAGAAAGGCTGAAGATCTGGAA 179
Db 51 SerGluArgLeuIysAspLeuGlu 58
|||||

RESULT 13
T29794
hypothetical protein T14G11.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29794
R:Fulton, B.; Hawkins, J.
submitted to the EMBL data Library, November 1995
A:Description: The sequence of *C. elegans* cosmid T14G11.
A:Reference number: Z20687
A:Accession: T29794
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-109 <FUL>
A:Cross-references: EMBL:U41024; PIDN:AAA82344.1; CESP:T14G11.2
C:Genetics:
A:Gene: CESP:T14G11.2
A:Introns: 11/3; 63/2

Alignment Scores:
Pred. No.: 85.5 Length: 109
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x T29794 (1-109)
QY 3005 GCCTCTGCACATGTAATCCACTA 3028
Db 61 AlaLeuCysThrLeuAsnProLeu 68
|||||

RESULT 14
B82271
hypothetical protein VC0862 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82271
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82271
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <HEI>
A:Cross-references: GB:AE004171; GB:AE003852; NID:g9955308; PIDN:AAF94024.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0862
A:Map position: 1

Alignment Scores:
Pred. No.: 84.8 Length: 122
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x B82271 (1-122)
QY 1094 ACTCAGAAGCGCACCTAACAGG 1117
Db 102 ThrGlnLysAlaThrLeuThrArg 109
|||||

RESULT 15
D71106
hypothetical protein PH0621 - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: D71106

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 17:05:34 ; Search time 27 Seconds

(without alignments)
12175.607 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atgcaagacagacagcataga.....tggagagtgtgaaacttttaa 3963

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09818990/runat_26112002_093221_22032/app_query.fasta_1.4103
-DB=SwissProt_40 -QMT=fastan -SUFFIX=oligo.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cd1
-LIST=1000 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFT=ptc -NORM=ext -HEAPS=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09818990 @CGN_1_1_26 -runat_26112002_093221_22032 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	10	0.8	804	1	MEAG_HUMAN
2	9	0.7	393	1	HVC3_HETFR
3	9	0.7	438	1	HVC2_HETFR
4	9	0.7	438	1	HVC5_HETFR
5	9	0.7	461	1	HVCN_HETFR
6	8	0.6	130	1	UL33_HSV11
7	8	0.6	227	1	Y006_DEIRA
8	8	0.6	302	1	ALSR_BACSU
9	8	0.6	312	1	CORA_WYCTU
10	8	0.6	323	1	HXDB_MOUSE
11	8	0.6	343	1	NOD1_RHISN
12	8	0.6	369	1	ODBB_RAT
13	8	0.6	374	1	PEL2_ERWCA
14	8	0.6	378	1	SLG1_YEAST
15	8	0.6	379	1	SGRA_METEX
16	8	0.6	382	1	YMF4_CAEEL
17	8	0.6	390	1	MOEA_SYNY3
18	8	0.6	392	1	ODBB_BOVIN

392	1	ODBB_HUMAN	P21953	homo sapien
401	1	CARA_AGR5	O8uaf7	agrobacteri
453	1	YD14_SCHPO	Q92342	schizosacch
487	1	CATA_STRCO	Q92398	streptomyce
504	1	AINX_MOUSE	P46660	mus musculus
516	1	GUX1_PHACH	P13860	phanerocha
557	1	APG6_YEAST	Q02948	saccharomyc
603	1	REF1_DROME	Q24492	drosophila
624	1	YBY3_SCHPO	Q10198	schizosacch
629	1	SKIL_MOUSE	Q60665	mus musculus
746	1	EXT1_CRIGR	O9jk82	cricketul
746	1	EXT1_HUMAN	Q16394	homo sapien
746	1	EXT1_MOUSE	P97464	mus musculus
1066	1	ITA3_CRISP	P17852	cricketdae
1095	1	PIPA_DROME	P13217	drosophila
1286	1	CTRO_HUMAN	O14578	homo sapien
1337	1	PTPJ_HUMAN	Q12913	homo sapien
1527	1	MRP3_HUMAN	O15438	homo sapien
1597	1	CTRO_MOUSE	P49025	mus musculus
1914	1	KMLS_HUMAN	O15746	homo sapien
2547	1	FAFX_HUMAN	Q93008	h probab
2559	1	FAFX_MOUSE	P70398	m probab
2747	1	FAF_DROME	P55824	drosophila
3726	1	ABF1_MOUSE	G61329	mus musculus
45	1	LYS3_SHISO	P21185	shigella so
49	1	LYS0_ECOLI	P02987	escherichia
52	1	LYS4_ECOLI	P09181	escherichia
72	1	PSBH_TOBAC	P06415	nicotiana t
87	1	Y476_MYCLE	O9cb43	mycobacteri
93	1	VG28_HSVSA	Q01009	herpesvirus
94	1	AFP_ASPCI	P17737	aspergillus
113	1	PSAK_CHLRE	P14225	chlamydomon
115	1	RK22_ODOSI	P49558	odontella s
118	1	ACHE_CHICK	P16005	gallus gall
118	1	YNIF_AZOBR	P25316	azospirillu
121	1	TKNK_HUMAN	Q9uhf0	homo sapien
121	1	Y733_METJA	O58143	methanococ
126	1	VAL5_ICMV	O8592	indian cass
130	1	RS9_LACIA	Q9c97	lactococcus
136	1	FLGE_YEREN	Q56893	yersinia en
136	1	NB8M_BOVIN	Q02368	bos taurus
139	1	INLA_HUMAN	Q14641	homo sapien
139	1	UMUD_SALTY	P22493	salmonella
144	1	SSI_STRAO	P21006	streptomyce
144	1	YA32_TREPA	O83995	treponema p
146	1	AROF_STRLI	P55911	streptomyce
152	1	SPI8_ARATH	O64644	arabidopsis
153	1	RL22_MYCSM	O06115	mycobacteri
154	1	YF12_MYCPN	P75274	mycoplasma
156	1	VE6_HPV41	P27555	human papil
157	1	Y012_BPL2	P42547	bacterioph
159	1	RK35_SPIOL	P23326	spinacia ol
161	1	Y4LG_RHISN	P55347	rhizobium s
173	1	GLBC_NIPBR	P51536	nippostrong
174	1	GRPE_METTH	O27350	methanobact
175	1	RL22_MYCLE	O32986	mycobacteri
177	1	CYSC_SYNY3	P72940	synecocyst
177	1	NR13_COTJA	Q90343	coturnix co
183	1	ZEB2_MAIZE	P08031	zea mays (m
187	1	PABA_KLEAE	P06194	klebsiella
190	1	CAH2_FLALI	P46513	flaveria li
190	1	PGHD_HUMAN	P41222	homo sapien
193	1	HUNE_DROIK	O46242	drosophila
195	1	RM32_DROME	Q9v921	drosophila
197	1	RL22_MYCTU	P95054	mycobacteri
198	1	FCPB_PHATR	Q08585	phaeodactyl
198	1	Y660_RITCCN	Q92hw0	rickettsia
201	1	YR21_TRSRV	P25245	tomato ring
206	1	YNJF_ECOLI	P76226	escherichia
214	1	YA64_METJA	Q58464	methanococ
215	1	YN68_DEIRA	Q9rrw7	deinococcc
216	1	HIS5_BRUME	Q8ye35	bruceella me
216	1	HIS5_RHIME	Q92tbl	rhizobium m

c 238	7	0.5	351	1	LPXD_BRUME	Q44630 bruceella me	c 311	7	0.5	427	1	IRF3_HUMAN	Q14653 homo sapien
c 239	7	0.5	352	1	CCR4_CERTO	O62747 cercocobus	312	7	0.5	427	1	MURE_CAMJE	O69290 campylobact
c 240	7	0.5	352	1	CCR4_HUMAN	P30991 homo sapien	313	7	0.5	428	1	TERM_DROME	P11455 drosophila
c 241	7	0.5	352	1	CCR4_MACFA	Q28474 macaca fasc	314	7	0.5	429	1	ARSB_ECOLI	P37310 escherichia
c 242	7	0.5	352	1	CCR4_WACMU	P75934 macaca mula	315	7	0.5	430	1	MENF_HAEIN	P44613 haemophilus
c 243	7	0.5	352	1	CCR4_PAPAN	P56491 papio anubi	316	7	0.5	431	1	HXB3_HUMAN	P14651 homo sapien
c 244	7	0.5	353	1	PRK1_SCHPO	O13958 schizosacch	317	7	0.5	431	1	PUR8_THEMEA	Q9x010 human cytom
c 245	7	0.5	353	1	CCR4_FELCA	P56498 felis silve	318	7	0.5	431	1	UL61_HCMVA	P16818 human cytom
c 246	7	0.5	354	1	POL1_TRSVR	P29150 tomato ring	319	7	0.5	433	1	AFUR_EMENT	P52957 emericeila
c 247	7	0.5	355	1	DPSE_CHICK	P28683 gallus gall	320	7	0.5	433	1	HXB3_MOUSE	P09026 mus musculus
c 248	7	0.5	356	1	DHSO_HUMAN	Q00796 homo sapien	321	7	0.5	435	1	PEPC_LACLC	Q04723 lactococcus
c 249	7	0.5	358	1	GP20_HUMAN	P99678 homo sapien	322	7	0.5	436	1	PAFA_CAVPO	P70683 c platelet-
c 250	7	0.5	359	1	CCR4_MOUSE	P70658 m c-x-c che	323	7	0.5	437	1	NCCC_ALCXX	Q44584 alcaligenes
c 251	7	0.5	360	1	IHA_BOVIN	P07994 bos taurus	324	7	0.5	437	1	NU4M_ALBUC	P48914 albinaria c
c 252	7	0.5	361	1	IHA_TRIVU	O77755 trichosurus	325	7	0.5	440	1	DCDA_STRCO	Q92hbs streptomyce
c 253	7	0.5	364	1	IHA_PIG	P04087 sus scrofa	326	7	0.5	440	1	Y284_AQUAE	O66338 aquifex aeo
c 254	7	0.5	366	1	IHA_HUMAN	P05111 homo sapien	327	7	0.5	442	1	CYS4_DICDI	P54639 dictyostell
c 255	7	0.5	366	1	IHA_MOUSE	Q04997 mus musculus	328	7	0.5	444	1	ASSY_HAEIN	P44315 haemophilus
c 256	7	0.5	366	1	IHA_RAT	P17490 rattus norv	329	7	0.5	445	1	ASSY_PASMU	P57877 pasteurella
c 257	7	0.5	366	1	IHA_RAT	P17490 rattus norv	330	7	0.5	445	1	NU4M_DINSE	O79555 dinodon sem
c 258	7	0.5	367	1	IHA_HORSE	P55101 equus cabal	331	7	0.5	445	1	YAV5_XANCV	P19520 xanthomonas
c 259	7	0.5	367	1	YMP4_STRCO	P43167 streptomyce	332	7	0.5	446	1	CITN_KLEPN	P1502 klebsiella
c 260	7	0.5	369	1	DNAJ_HELPU	Q92792 helicobacte	333	7	0.5	446	1	CITN_SALDU	P31603 salmonella
c 261	7	0.5	369	1	DNAJ_HELPY	O25890 helicobacte	334	7	0.5	446	1	CITN_SALTY	P31604 salmonella
c 262	7	0.5	369	1	O24A_DROME	P81913 drosophila	335	7	0.5	448	1	G6PI_THEMEA	Q9x1a5 thermotoga
c 263	7	0.5	371	1	NUIM_NEUCR	P08774 neurospora	336	7	0.5	449	1	CLUS_HUMAN	P10909 homo sapien
c 264	7	0.5	374	1	CYS1_CAPAN	P31300 capsicum an	337	7	0.5	450	1	PAX8_HUMAN	O06710 homo sapien
c 265	7	0.5	375	1	DP3B_SYN7P	P52023 synechococc	338	7	0.5	456	1	TRME_BUCAP	Q44633 buchnera ap
c 266	7	0.5	375	1	OE56_NPVBM	Q92500 bombyx mori	339	7	0.5	456	1	TRME_LEPIN	P97043 leptospira
c 267	7	0.5	375	1	TRMU_CAEEL	Q17440 caenorhabdi	340	7	0.5	457	1	PAX8_MOUSE	Q00288 mus musculus
c 268	7	0.5	377	1	MTB1_BRUAB	Q03570 bruceella ab	341	7	0.5	458	1	PAX8_RAT	P51974 rattus norv
c 269	7	0.5	377	1	VE2_HPV44	Q80916 human papil	342	7	0.5	459	1	NCB1_MOUSE	Q02819 mus musculus
c 270	7	0.5	378	1	MUTY_HAEIN	P44320 haemophilus	343	7	0.5	459	1	NCB1_RAT	Q63083 rattus norv
c 271	7	0.5	380	1	BIOF_HAEIN	P44422 haemophilus	344	7	0.5	459	1	NU4M_POLOR	Q95917 polypterus
c 272	7	0.5	380	1	COSC_YEAST	P53053 saccharomyc	345	7	0.5	459	1	PAX8_CANFA	P47240 canis famil
c 273	7	0.5	380	1	PKN5_MYXXA	P54737 myxococcus	346	7	0.5	460	1	NU4M_ASTPE	P11992 asterina pe
c 274	7	0.5	380	1	YMH7_CAEEL	P34474 caenorhabdi	347	7	0.5	460	1	NU4M_BRARE	Q9m1y1 brachydanio
c 275	7	0.5	382	1	DHGY_CUCSA	P13443 cucumis sat	348	7	0.5	460	1	NU4M_CARAU	O78687 carassius a
c 276	7	0.5	382	1	DXY_BACHD	Q9ka69 bacillus ba	349	7	0.5	460	1	NU4M_CROLA	P34194 crossostoma
c 277	7	0.5	384	1	SSR4_RAT	P30937 rattus norv	350	7	0.5	460	1	NU4M_GADMO	P55781 gadus morhu
c 278	7	0.5	388	1	SSR4_HUMAN	P31391 homo sapien	351	7	0.5	460	1	NU4M_ONCMY	P11631 oncorhynch
c 279	7	0.5	392	1	SPYA_HUMAN	P21549 homo sapien	352	7	0.5	460	1	NU4M_SALSA	Q9zxm4 salmo salar
c 280	7	0.5	392	1	SPYA_RABIT	P31030 cryptotlagus	353	7	0.5	460	1	NU4M_SCYCA	O79410 scyllorhinu
c 281	7	0.5	396	1	RFIM_SCHPO	Q09691 schizosacch	354	7	0.5	460	1	NU4M_SQUAC	Q9z245 squallus aca
c 282	7	0.5	397	1	ALR1_RHIL0	O98a05 rhizobium l	355	7	0.5	461	1	MNT2_CANAL	P46592 candida alb
c 283	7	0.5	399	1	Y507_PASMU	O9enc5 pasteurella	356	7	0.5	461	1	NCB1_HUMAN	Q02818 homo sapien
c 284	7	0.5	401	1	LSG1_HAEIN	P71399 haemophilus	357	7	0.5	461	1	NU4M_LATCH	Q03173 latimeria c
c 285	7	0.5	402	1	YIN2_STRAM	P32426 streptomyce	358	7	0.5	462	1	OCT2_MOUSE	Q00196 mus musculus
c 286	7	0.5	404	1	Y4C1_RHISN	P55391 rhizobium s	359	7	0.5	465	1	SPE4_CAEEL	Q01608 caenorhabdi
c 287	7	0.5	404	1	Y867_HAEIN	O57484 haemophilus	360	7	0.5	470	1	EGR2_RAT	P51774 rattus norv
c 288	7	0.5	404	1	ZFS1_SCHPO	P47979 schizosacch	361	7	0.5	471	1	PD22_DROVI	Q24705 drosophila
c 289	7	0.5	405	1	Y57A_HUMAN	O60320 homo sapien	362	7	0.5	472	1	POU2_BRARE	Q90270 brachydanio
c 290	7	0.5	405	1	YRKO_BACSU	P54442 bacillus su	363	7	0.5	474	1	LCPT_STRIN	O33654 streptococ
c 291	7	0.5	406	1	IPSP_HUMAN	P05154 homo sapien	364	7	0.5	474	1	VTDB_HUMAN	P02774 homo sapien
c 292	7	0.5	409	1	ARCA_LACSK	O53088 lactobacill	365	7	0.5	475	1	YDF6_SCHPO	Q10478 schizosacch
c 293	7	0.5	409	1	OX12_SCHPO	O43092 schizosacch	366	7	0.5	476	1	PURA_WHEAT	O24396 triticum ae
c 294	7	0.5	410	1	ARCA_LACLA	P58013 lactococcus	367	7	0.5	478	1	OCT2_HUMAN	P09086 homo sapien
c 295	7	0.5	410	1	VE2_BPVI	P03122 bovine papi	368	7	0.5	478	1	OCT2_PIG	Q29013 sus scrofa
c 296	7	0.5	411	1	IF2G_PYRAB	Q9v190 pyrococcus	369	7	0.5	481	1	PRTB_ERWCH	P16316 erwinia chr
c 297	7	0.5	411	1	IF2G_PYRFO	O8u082 pyrococcus	370	7	0.5	484	1	IRX5_MOUSE	Q9jfk4 mus musculus
c 298	7	0.5	411	1	IF2G_PYRHO	O59410 pyrococcus	371	7	0.5	487	1	TTDT_ECOLI	P39414 escherichia
c 299	7	0.5	411	1	VE2_BPVI	P11299 bovine papi	372	7	0.5	488	1	HH1R_CAVPO	P31389 cavia porce
c 300	7	0.5	413	1	ARCA_BACLI	O86131 bacillus li	373	7	0.5	490	1	ILVC_ECO57	P58256 escherichia
c 301	7	0.5	413	1	ARCA_CLOPE	Q46254 clostridium	374	7	0.5	490	1	ILVC_ECOLI	P05793 escherichia
c 302	7	0.5	414	1	TYT1_HUMAN	P25490 homo sapien	375	7	0.5	490	1	ILVC_SALTY	P05989 salmonella
c 303	7	0.5	414	1	TYT1_MOUSE	Q00899 mus musculus	376	7	0.5	490	1	SUCP_LEUME	O59495 leuconostoc
c 304	7	0.5	416	1	MTPI_PSYTA	O33481 psychobact	377	7	0.5	491	1	VIEL_HCMVA	P13202 human cytom
c 305	7	0.5	417	1	CZCC_ALCEU	P13509 alcaligenes	378	7	0.5	493	1	NUBM_NEUCR	P24917 neurospora
c 306	7	0.5	417	1	IRX5_HUMAN	P78411 homo sapien	379	7	0.5	495	1	ENP2_RAT	O35795 rattus norv
c 307	7	0.5	419	1	K1C4_XENLA	P05781 xenopus lae	380	7	0.5	495	1	VLL_BPVI	P03103 bovine papi
c 308	7	0.5	422	1	OAT_CAEEL	Q18040 caenorhabdi	381	7	0.5	496	1	NUBM_ASPNG	O92406 aspergillus
c 309	7	0.5	425	1	Y450_HUMAN	O75038 homo sapien	382	7	0.5	496	1	SUMT_SCHPO	O74468 schizosacch
c 310	7	0.5	425	1	YNN6_YEAST	P53911 saccharomyc	383	7	0.5	497	1	VLL_BPVI	P06458 bovine papi

c 822	100	0.5	6	895	6	0.5	112	1	CYC_CRION	P00077 crithidia o
823	100	0.5	6	c 896	6	0.5	112	1	MYRL_MYRPI	Q07932 myrmecia pi
824	100	0.5	6	c 897	6	0.5	112	1	NAPD_PARP	Q05349 paracoccus
c 825	101	0.5	6	c 898	6	0.5	112	1	PSC2_RAT	P02781 rattus norv
c 826	101	0.5	6	c 899	6	0.5	112	1	RE30_EUPES	Q9m36 euphorbia e
827	101	0.5	6	c 900	6	0.5	112	1	VL4_BP2	P11301 bovine papi
828	101	0.5	6	c 901	6	0.5	113	1	DAD1_HUMAN	P46966 homo sapien
829	101	0.5	6	902	6	0.5	113	1	DAD1_PIG	Q29036 sus scrofa
830	102	0.5	6	903	6	0.5	113	1	YG35_BPMV4	Q04775 lactococcus
c 831	102	0.5	6	c 904	6	0.5	113	1	YGAM_ECOLI	Q04713 escherichia
c 832	102	0.5	6	905	6	0.5	114	1	CPCL_CANPG	P81380 cancer pagu
c 833	102	0.5	6	906	6	0.5	114	1	H37_STYLE	P81201 stylynychia
c 834	102	0.5	6	907	6	0.5	114	1	H37_STYLE	P81196 stylynychia
835	103	0.5	6	908	6	0.5	114	1	HYPA_CAMJE	Q9p90 campylobact
836	103	0.5	6	c 909	6	0.5	114	1	KL16_RABIT	P01697 oryctolagus
837	103	0.5	6	c 910	6	0.5	114	1	NL22_TOBAC	Q03461 nicotiana t
838	103	0.5	6	c 911	6	0.5	114	1	UL96_FSW7J	P52475 human herpe
839	104	0.5	6	c 912	6	0.5	114	1	V103_FOPPV	Q9j57 fowlpox vir
c 840	104	0.5	6	c 913	6	0.5	115	1	ACPS_UREPA	Q9p97 ureaplasma
841	104	0.5	6	c 914	6	0.5	115	1	NU3M_HYLLA	Q95708 hyllobates l
842	104	0.5	6	c 915	6	0.5	115	1	NUFM_RAT	Q63362 rattus norv
843	105	0.5	6	916	6	0.5	115	1	REV_HV1MN	P05871 human immun
c 844	105	0.5	6	917	6	0.5	115	1	SMSI_PROAN	Q9w10 protopterus
c 845	105	0.5	6	c 918	6	0.5	115	1	TAT_SIVGB	P22384 simian immu
c 846	105	0.5	6	c 919	6	0.5	115	1	YEAQ_ECOLI	P76243 escherichia
847	105	0.5	6	920	6	0.5	116	1	GLHA_CLAGA	P53542 clarias gar
848	105	0.5	6	921	6	0.5	116	1	GLHA_ICTPU	Q9y9p3 ictalurus p
c 849	105	0.5	6	c 922	6	0.5	116	1	NLTP_CICAR	Q23758 cicier ariet
c 850	105	0.5	6	923	6	0.5	116	1	NU3M_ONCKE	P20686 oncorhynch
c 851	105	0.5	6	924	6	0.5	116	1	NU3M_ONCKE	Q35262 oncorhynch
c 852	105	0.5	6	925	6	0.5	116	1	NU3M_ONCKI	P20687 oncorhynch
c 853	106	0.5	6	926	6	0.5	116	1	NU3M_ONCKA	Q37108 oncorhynch
c 854	106	0.5	6	927	6	0.5	116	1	NU3M_ONCMY	P11629 oncorhynch
c 855	106	0.5	6	928	6	0.5	116	1	NU3M_ONCNE	P20688 oncorhynch
c 856	106	0.5	6	929	6	0.5	116	1	NU3M_ONCTS	P25707 oncorhynch
857	106	0.5	6	930	6	0.5	116	1	NU3M_SALSA	Q35929 salmo salar
c 858	107	0.5	6	931	6	0.5	116	1	NU3M_SALTR	Q03252 salmo trutt
c 859	107	0.5	6	c 932	6	0.5	116	1	REV_HV1S1	P19548 human immun
860	107	0.5	6	933	6	0.5	116	1	REV_HV1Y2	P35960 human immun
c 861	107	0.5	6	c 934	6	0.5	116	1	RL19_FUSNN	Q8r980 fusobacteri
c 862	107	0.5	6	c 935	6	0.5	116	1	TAT_SIVAM	P36340 simian immu
c 863	107	0.5	6	c 936	6	0.5	116	1	Y105_ARATH	Q9x1k3 arabidopsis
c 864	107	0.5	6	937	6	0.5	116	1	YE94_HAEIN	Q05071 haemophilus
c 865	107	0.5	6	938	6	0.5	116	1	YG1Q_YEAST	P53222 saccharomyc
c 866	108	0.5	6	939	6	0.5	117	1	HV02_CAICR	P03981 canian croc
c 867	108	0.5	6	940	6	0.5	117	1	HV02_CANFA	P01785 canis famil
868	108	0.5	6	941	6	0.5	117	1	HV03_CANAU	P19180 carassius a
869	108	0.5	6	c 942	6	0.5	117	1	HV05_MOUSE	P01749 mus musculu
870	108	0.5	6	c 943	6	0.5	117	1	HV06_MOUSE	P01750 mus musculu
c 871	108	0.5	6	944	6	0.5	117	1	R18E_ARCFU	Q29138 archaeoglob
c 872	108	0.5	6	945	6	0.5	117	1	RBS1_CHRFU	P22850 chromatium
c 873	108	0.5	6	946	6	0.5	117	1	TLB2_MOUSE	P56841 mus musculu
c 874	108	0.5	6	947	6	0.5	118	1	GLH1_CYPCA	P01221 cyprinus ca
c 875	109	0.5	6	948	6	0.5	118	1	GLHA_CTEID	P30983 ctenopharyn
c 876	109	0.5	6	949	6	0.5	118	1	GLHA_HYPMO	P37037 hypophthalm
c 877	109	0.5	6	c 950	6	0.5	118	1	PA2H_LATCO	P10117 latitcauda c
c 878	109	0.5	6	951	6	0.5	118	1	SYC_STAXY	P77986 staphylococ
c 879	109	0.5	6	952	6	0.5	118	1	TRH4_ECOLI	Q00190 escherichia
c 880	109	0.5	6	953	6	0.5	118	1	VATF_YEAST	P39111 saccharomyc
c 881	110	0.5	6	954	6	0.5	119	1	B2MG_MOUSE	P01887 mus musculu
c 882	110	0.5	6	955	6	0.5	119	1	RBFA_LACLA	Q9ch3 lactococcus
c 883	110	0.5	6	c 956	6	0.5	119	1	YEAR_ECOLI	P76248 escherichia
c 884	110	0.5	6	c 957	6	0.5	119	1	YOAC_ECOLI	P76259 escherichia
c 885	110	0.5	6	958	6	0.5	119	1	YPAA_LACLA	Q02009 lactococcus
886	110	0.5	6	959	6	0.5	120	1	NU3C_MESVI	Q9mud9 mesostigma
887	110	0.5	6	960	6	0.5	120	1	PFDB_METKA	Q8tyc7 methanopyru
c 888	111	0.5	6	c 961	6	0.5	120	1	VE4_BPVI	P06926 bovine papi
c 889	111	0.5	6	962	6	0.5	121	1	ANRE_HUMAN	P15267 homo sapien
890	111	0.5	6	c 963	6	0.5	121	1	NU3M_PORPU	Q99978 porphyra pu
891	111	0.5	6	964	6	0.5	121	1	RNPA_CORBU	P45648 coxella bu
892	111	0.5	6	965	6	0.5	121	1	VG38_HSVII	Q00144 ictaluriid h
893	111	0.5	6	c 966	6	0.5	121	1	YEDR_ECOLI	P76334 escherichia
894	112	0.5	6	967	6	0.5	122	1	MP13_LYMST	P80090 lymnaea sta

```
968 6 0.5 122 1 YRHF_BACSU
c 969 6 0.5 123 1 C59A_MOUSE
c 970 6 0.5 123 1 CRCB_PYRHO
971 6 0.5 123 1 ELI2_PHYCR
972 6 0.5 123 1 ELI3_PHYCR
c 973 6 0.5 123 1 PFDB_AERPE
c 974 6 0.5 123 1 RBFA_RALSO
975 6 0.5 123 1 RBFA_SPIOL
976 6 0.5 123 1 RBS1_NOC5I
977 6 0.5 124 1 SVLD_HUMAN
c 978 6 0.5 124 1 MAG5_HUMAN
c 979 6 0.5 124 1 RS16_RHIME
c 980 6 0.5 124 1 URE2_BACSU
981 6 0.5 124 1 VATE_NEUCR
c 982 6 0.5 124 1 VE4_HPV57
983 6 0.5 125 1 HV1F_HUMAN
c 984 6 0.5 125 1 RS13_RICPR
c 985 6 0.5 126 1 CCB3_ECOLI
c 986 6 0.5 126 1 CCB4_ECOLI
c 987 6 0.5 126 1 CLP2_DROME
c 988 6 0.5 126 1 PRO1_DICDI
c 989 6 0.5 126 1 RBS1_ACECL
990 6 0.5 126 1 SECE_VIBCH
991 6 0.5 126 1 VXS1_BPAPS
c 992 6 0.5 127 1 CY3A_DESAF
993 6 0.5 127 1 GTH1_ANGJA
c 994 6 0.5 127 1 LY6D_MOUSE
c 995 6 0.5 127 1 RL17_XANCP
c 996 6 0.5 127 1 YLIH_ECOLI
c 997 6 0.5 128 1 CION_ECIOIN
c 998 6 0.5 128 1 RNF18_MAIZE
999 6 0.5 128 1 RNP_PREEN
1000 6 0.5 128 1 YBM6_YEAST
```

ALIGNMENTS

```
RESULT 1
MEAG6_HUMAN STANDARD; PRT; 804 AA.
AC Q15320: Q00169;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Meningioma-expressed antigen 6/11 (MEAG6) (MEAL1).
GN MGEAG6 OR MGEAL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS MEAG6 AND MEAL1).
RP TISSUE=Meningioma;
RX MEDLINE=97472454; PubMed=9356211;
RA Heckel D., Brass N., Fischer U., Blin N., Steudel I., Tuercei O.,
RA Fackler O., Zang K.D., Meese E.;
RT *cDNA cloning and chromosomal mapping of a predicted coiled-coil
RT proline-rich protein immunogenic in meningioma patients.";
RL Hum. Mol. Genet. 6:2031-2041(1997).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; MEAG6 (SHOWN HERE) AND MEAL1; MAY
CC BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED IN BRAIN, MUSCLE
CC AND CRANIAL SKIN.
CC -1- DISEASE: AUTOANTIGEN FOUND IN BENIGN MENINGIOMA.
CC -1- CAUTION: AT LEAST 9 PSEUDOGENES HAVE BEEN FOUND ON DIFFERENT
CC CHROMOSOMES, INCLUDING CHROMOSOMES 2, 3, 6, 7, 9, 10, 12, 13 AND
CC 18.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X07782; CAA30615.1; -
CC PIR; S01852; HVRKCL.
CC HSSP; P01842; 7FAB.
CC InterPro: IPR003006; Iq_MHC.
CC InterPro: IPR003597; Iq_cl.
CC InterPro: IPR003600; Iq_like.
CC Pfam; PF00047; Iq; 3.
CC SMART; SM00410; Iq_like; 1.
-----
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U94780; AAB86593.1; -
EMBL; U73682; AAB86589.1; ALT_INIT.
DR Genew; HGNC:7057; MGEAG6.
DR MIM; 602132; -
KW Antigen; Coiled coil; Alternative splicing.
FT DOMAIN 87 273 COILED COIL (POTENTIAL).
FT DOMAIN 321 501 COILED COIL (POTENTIAL).
FT DOMAIN 553 558 POLY-GLY.
FT DOMAIN 502 804 PRO-RICH.
FT VARSPPLIC 514 556 MISSING (IN ISOFORM MEAL1).
FT CONFLICT 198 198 P -> O (IN REF. 1; AAB86589).
FT CONFLICT 360 360 E -> O (IN REF. 1; AAB86589).
FT CONFLICT 589 589 L -> P (IN REF. 1; AAB86589).
FT CONFLICT 594 594 F -> S (IN REF. 1; AAB86589).
FT CONFLICT 699 699 V -> I (IN REF. 1; AAB86589).
FT CONFLICT 738 738 G -> R (IN REF. 1; AAB86589).
SQ SEQUENCE 804 AA; 90943 MW; 3F088799E0883EDE CRC64;
Alignment Scores:
Pred. No.: 0.293 Length: 804
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.76% Indels: 0
DB: 1 Gaps: 0
US-09-818-990b-1 (1-3963) x MEAG6_HUMAN (1-804)
QY 161 GCGGAGGCCAAGATGACCTCCAGATCTTT 190
| | | | | | | | | | | | | | | | | | | |
Db 188 AlaGluAlaLysMetThrPheGlnIlePhe 197
RESULT 2
HVC3_HETFR STANDARD; PRT; 393 AA.
ID HVC3_HETFR STANDARD; PRT; 393 AA.
AC P23086;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region (Clone 6121) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Spleen;
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Litman R., Shambloott M.J., Litman G.W.;
RT *Complete structure and organization of immunoglobulin heavy chain
RT constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X07782; CAA30615.1; -
CC PIR; S01852; HVRKCL.
CC HSSP; P01842; 7FAB.
CC InterPro: IPR003006; Iq_MHC.
CC InterPro: IPR003597; Iq_cl.
CC InterPro: IPR003600; Iq_like.
CC Pfam; PF00047; Iq; 3.
CC SMART; SM00410; Iq_like; 1.
```

```
DR SMART; SM00407; IGcl; 2.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 1
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 393 AA; 43081 MW; 4E44D076972F18B5 CRC64;

Alignment Scores:
Pred. No.: 3 37 Length: 393
Score: 9.00 Matches: 9
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.68%
Indels: 0
DB: 1 0

US-09-818-990B-1 (1-393) x HVC3_HETFR (1-393)
QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706
| | | | | | | | | | | | | | | | | | | |
Db 251 LeuProThrProLeuArgAlaSerIle 259

RESULT 3
HVC2_HETFR
ID HVC2_HETFR STANDARD; PRT; 438 AA.
AC P23085;
DT 01-NOV-1991 (Rel. 20, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Ig heavy chain C region (Clone 12022) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Litman R., Shambloott M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07784; CAA30617.1; -.
CC PIR; S00980; HVRKC2.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR003597; Ig_cl.
CC InterPro; IPR003600; Ig_like.
CC Pfam; PF00047; Ig; 4.
CC SMART; SM00410; IG_like; 2.
CC SMART; SM00407; IGcl; 2.
CC PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 1
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 438 AA; 47904 MW; 6E67A9C5BBF7AA7B CRC64;

Alignment Scores:
Pred. No.: 3 35 Length: 438
Score: 9.00 Matches: 9
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 0.68%
Indels: 0
DB: 1 0

US-09-818-990B-1 (1-393) x HVC2_HETFR (1-438)
QY 1680 CTCGCCACCCACTCAGAGCTCCATC 1706
| | | | | | | | | | | | | | | | | | | |
Db 296 LeuProThrProLeuArgAlaSerIle 304

RESULT 4
HVC2_HETFR
ID HVC2_HETFR STANDARD; PRT; 438 AA.
AC P23087;
DT 01-NOV-1991 (Rel. 20, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Ig heavy chain C region, secreted form (Clone 3050).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinds K., Litman R., Shambloott M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07781; CAA30613.1; ALT_INIT.
CC PIR; S01853; HVRKCS.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR003597; Ig_cl.
CC InterPro; IPR003600; Ig_like.
CC Pfam; PF00047; Ig; 4.
CC SMART; SM00410; IG_like; 2.
CC SMART; SM00407; IGcl; 2.
CC PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing.
FT NON_TER 1 1
FT DOMAIN 1 99 CH1.
FT DOMAIN 100 205 CH2.
FT DOMAIN 206 308 CH3.
FT DOMAIN 309 418 CH4.
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
```


DR InterPro: IPR005119; LysR_subst.
 DR Pfam: PF00126; HTH_1; 1.
 DR Pfam: PF03466; LysR_substrate; 1.
 DR PRINTS: PR00039; HTHLYSR.
 DR PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
 KW Transcription regulation; DNA-binding; Acetoin biosynthesis;
 KW Complete proteome.
 FT DNA_BIND 18 37 H-T-H MOTIF (BY SIMILARITY).
 SQ SEQUENCE 302 AA; 34333 MW; BCA4DF5CF95A3E33 CRC64;

Alignment Scores:
 Pred. No.: 37.7 Length: 302
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x ALSR_BACSU (1-302)

QY 659 CTGCTATCCGAGGATAGGATG 636
 |||||
 Db 232 LeuValSerAlaGlyIleGlyMet 239

RESULT 9

COAA_MYCTU
 ID COAA_MYCTU STANDARD; PRT; 312 AA.
 AC O53440;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pantothenate kinase (EC 2.7.1.33) (Pantothenic acid kinase).
 GN COAA OR RV1092C OR MT1124 OR MTV017.45C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-
 phosphopantothenate.
 CC -1- PATHWAY: Coenzyme A (CoA) biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC -1- SIMILARITY: BELONGS TO THE PANTOTHENATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AL021897; CAAL7208.1; -
 DR EMBL: AE006992; AAK45382.1; -
 DR TIGR: MT1124; -
 DR TubercuList; Rv1092c; -
 DR InterPro: IPR001324; PRK.
 DR InterPro: IPR004566; Pank_bact.
 DR Pfam: PF00485; PRK; 1.
 DR TIGRFAMS: TIGR00554; pank_bact; 1.
 KW Transferase; Kinase; ATP-binding; Coenzyme A biosynthesis;
 KW Complete proteome.
 FT NP_BIND 97 104 ATP (POTENTIAL).
 SQ SEQUENCE 312 AA; 35656 MW; F075B543AE75788D CRC64;

Alignment Scores:
 Pred. No.: 37.6 Length: 312
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x COAA_MYCTU (1-312)

QY 383 GAGCTGGTGGACTTCGAGGGTTA 360
 |||||
 Db 33 GluLeuValGlyLeuArgGlyLeu 40

RESULT 10

HXDB_MOUSE
 ID HXDB_MOUSE STANDARD; PRT; 323 AA.
 AC P23813;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-d11 (Hox-4.6) (Hox-5.5).
 GN HOXD11 OR HOXD-11 OR HOX-4.6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91293104; PubMed=1676674;
 RA Izpisua-Belmonte J.-C., Falkenstein H., Dolle P., Renucci A.,
 Duboule D.;
 RT "Murine genes related to the Drosophila AbdB homeotic genes are
 sequentially expressed during development of the posterior part of
 the body."
 RL EMBO J. 10:2279-2289(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv;
 RX MEDLINE=94074553; PubMed=7902810;
 RA Gerard M., Duboule D., Zakany J.;
 RT "Structure and activity of regulatory elements involved in the
 activation of the Hoxd-11 gene during late gastrulation."
 RL EMBO J. 12:3539-3550(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Sv; TISSUE=Liver;
 RA Takada S., Cook M., Kramlauf R., McMahon A.P.;
 RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING DEVELOPMENT OF THE
 POSTERIOR PART OF THE BODY.
 CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEODOMAIN FAMILY.


```

-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X60761; CAA43173.1; -
DR EMBL; X60762; CAA43173.1; JOINED.
DR EMBL; X71422; CAA50553.1; -
DR EMBL; X60395; CAA42943.1; ALT_INIT.
DR PIR; S16318; S16318.
DR PIR; S40403; S40403.
DR HSSP; P14653; 1B72.
DR TRANSFAC; T01760; -
DR MGD; MGI:96203; Hoxd11.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 64 87 POLY-GLY.
FT DOMAIN 93 105 POLY-ALA.
FT DNA_BIND 251 310 HOMEBOX.
SQ SEQUENCE 323 AA; 33545 MW; B003D1347F5FE771 CRC64;

Alignment Scores:
Pred. No.: 37.5 Length: 323
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.61% Indels: 0
DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x HXDB_MOUSE (1-323)
QY 145 GGGCGCGCTGAGGAGCGGAGGC 168
|||||
DB 221 GlyAlaAlaGluGlyGlyGly 228

RESULT 11
NODI_RHISN STANDARD; PRT; 343 AA.
AC P55476;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nodulation ATP-binding protein I.
GN NODI OR Y4HF.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
-----

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE000076; AAB91694.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR TIGRFAMS; TIGR01288; nodi; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Nodulation; Membrane; ATP-binding; Transport; Plasmid.
FT NP_BIND 77 84 ATP (POTENTIAL).
SQ SEQUENCE 343 AA; 37917 MW; F49A7EC56E099A33 CRC64;

Alignment Scores:
Pred. No.: 37.4 Length: 343
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x NODI_RHISN (1-343)
QY 169 GGCCTCCGCTCTTCAGCGGCC 146
|||||
DB 319 GlyLeuArgLeuLeuGlnArgPro 326

RESULT 12
ODBB_RAT STANDARD; PRT; 369 AA.
AC P35738;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 2-oxoisovalerate dehydrogenase beta subunit, mitochondrial precursor
DE (EC 1.2.4.4) (Branched-chain alpha-keto acid dehydrogenase component
DE beta chain (E1)) (BCKDH E1-beta) (Fragment).
GN BCKDHB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-54.
RX MEDLINE=93003326; PubMed=1390893;
RA Zhao Y., Kuntz M.J., Harris R.A., Crabb D.W.;
RT "Molecular cloning of the E1 beta subunit of the rat branched chain
RT alpha-ketoacid dehydrogenase.";
RL Biochim. Biophys. Acta 1132:207-210(1992).
RN [2]
RP ERRATUM.
RX MEDLINE=95143286; PubMed=7841205;
RA Zhao Y., Kuntz M.J., Harris R.A., Crabb D.W.;
RL Biochim. Biophys. Acta 1260:243-243(1995).
CC -!- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS:
CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE
CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).
CC -!- CATALYTIC ACTIVITY: 3-methyl-2-oxobutanoate + lipoamide -> S-(2-
CC methylpropanoyl)dihydroliipoamide + CO(2).
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
-----

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M94040; AAA73899.1; -;
 DR PIR; S28950; S28950.
 DR HSSP; P09061; IQSO.
 DR InterPro; IPR000360; Transketolase.
 DR Pfam; PF02779; transket_pyr; 1.
 DR Pfam; PF02780; transketolase_C; 1.
 KW Oxidoreductase; Mitochondrion; Transit peptide.
 FT NON_TER 1
 FT TRANSIT <1 27 MITOCHONDRION.
 FT CHAIN 28 369 2-OXOISOMALATE DEHYDROGENASE BETA
 FT SUBUNIT
 SQ SEQUENCE 369 AA; 40561 MW; 33E7B84EAB8178FD CRC64;

Alignment Scores:
 Pred. No.: 37.2 Length: 369
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x ODBB_RAT (1-369)

QY 2289 GAGGCTTCTTGTCTACTGCTCT 2266

DB 178 GLUALAPheAlaHisCysPro 185
 |||||

RESULT 13

ID PEL2_ERWCA STANDARD; PRT; 374 AA.
 AC P11431; Q06112; Q47469;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Pectate lyase II precursor (EC 4.2.2.2) (PEL II) (PLB).
 GN PEL2 OR PELB.
 OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 ON NCBI_TaxID=554;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ER / IAM1068;
 RX MEDLINE=93113068; PubMed=1369060;
 RA Yoshida A., Matsuo Y., Kamio Y., Izaki K.;
 RT "Molecular cloning and sequencing of the extracellular pectate lyase
 RT II gene from Erwinia carotovora Er";
 RL Biosci. Biotechnol. Biochem. 56:1596-1600(1992).
 RN [2]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=Sp. atroseptica EC;
 RX MEDLINE=87308030; PubMed=3040692;
 RA Lei S.-P., Lin H.-C., Wang S.-S., Callaway J., Wilcox G.;
 RT "Characterization of the Erwinia carotovora pelB gene and its product
 RT pectate lyase";
 RL J. Bacteriol. 169:4379-4383(1987).
 RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Sp. atroseptica C18.
 RX MEDLINE=95291436; PubMed=7773390;
 RA Bartling S., Wegener C., Olsen O.;
 RT "Synergism between Erwinia pectate lyase isoenzymes that depolymerize
 RT both pectate and pectin";
 RL Microbiology 141:873-881(1995).
 CC -1- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
 CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at
 CC their non-reducing ends.

CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1. PLADES
 CC SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; S51475; AAC60422.1; -;
 DR EMBL; M17364; AAA24848.1; -;
 DR EMBL; X81847; CAA57440.1; -;
 DR HSSP; P11073; IAIR.
 DR InterPro; IPR002022; Amb_allergen.
 DR Pfam; PF00544; pec_lyase; 1.
 KW Lyase; Multigene family; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 374 PECTATE LYASE II.
 FT DISULFID 93 176 BY SIMILARITY.
 FT DISULFID 350 373 BY SIMILARITY.
 FT ACT_SITE 239 239 POTENTIAL.
 FT VARIANT 35 35 E -> D (IN STRAIN EC).
 FT VARIANT 47 48 MK -> LQ (IN STRAIN EC).
 FT VARIANT 58 58 Q -> K (IN STRAIN EC).
 FT VARIANT 78 78 S -> N (IN STRAIN EC).
 FT VARIANT 136 136 V -> L (IN STRAIN C18).
 FT VARIANT 139 139 M -> I (IN STRAIN C18).
 FT VARIANT 144 144 M -> I (IN STRAIN C18).
 FT VARIANT 150 150 D -> H (IN STRAIN C18).
 FT VARIANT 156 156 I -> V (IN STRAIN EC).
 FT VARIANT 168 168 E -> K (IN STRAIN EC).
 FT VARIANT 172 173 KN -> QS (IN STRAIN EC).
 FT VARIANT 234 234 R -> S (IN STRAIN C18).
 FT VARIANT 252 252 N -> T (IN STRAIN EC).
 FT VARIANT 280 280 N -> I (IN STRAIN EC).
 FT VARIANT 306 311 SPSPFA -> KPADES (IN STRAINS C18 AND EC).
 FT VARIANT 314 314 K -> N (IN STRAIN EC).
 FT VARIANT 319 319 K -> R (IN STRAIN EC).
 FT VARIANT 322 322 S -> T (IN STRAIN EC).
 FT VARIANT 325 325 I -> V (IN STRAIN EC).
 FT VARIANT 329 329 N -> D (IN STRAINS C18 AND EC).
 FT VARIANT 338 339 SI -> AV (IN STRAIN C18).
 SQ SEQUENCE 374 AA; 40380 MW; FAE3FD00ACD29ED5 CRC64;

Alignment Scores:

Pred. No.: 37.2 Length: 374
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.63% Indels: 0
 DB: 1 Gaps: 0

US-09-818-990b-1 (1-3963) x PEL2_ERWCA (1-374)

QY 3036 ATCACTGGTAGTCGATCAATGTC 3013

DB 258 IleThrGlySerGlyPheAsnVal 265
 |||||

RESULT 14

ID SLG1_YEAST STANDARD; PRT; 378 AA.
 AC P54867;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SLG1 protein precursor.
 GN SLG1 OR YOR008C OR UNP378.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bettignies G., Bergez-Aulio P., Barthe C., Louvet O.,
RA Peypouquet M.F., Morel C., Doignon F., Crouzet M.;
RL Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051599; PubMed=8896276;
RA Sterky F., Holmberg A., Pettersson B., Uhlen M.;
RT "The sequence of a 30 kb fragment on the left arm of chromosome XV
RT from Saccharomyces cerevisiae reveals 15 open reading frames, five of
RT which correspond to previously identified genes.";
RL Yeast 12:1091-1095(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39481; AAC85862.1; -
DR EMBL: U43491; AAC49488.1; -
DR EMBL: Z74916; CAA99196.1; -
DR SGD: S0005534; SLG1.
DR InterPro: IPR002889; WSC.
DR Pfam: PF01822; WSC; 1.
DR SMART: SM00321; WSC; 1.
DR Glycoprotein: Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 378 SLG1 PROTEIN.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 378 AA; 39270 MW; EEE164F2374CCCE3 CRC64;

Alignment Scores:
Pred. No.: 37.1 Length: 378
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x SLG1_YEAST (1-378)

QY 2795 ATTCATCATCTGATTCATCAACA 2772
|||||
DB 118 ILeSerSerSerSerSerSerThr 125

RESULT 15
SGAA_METEX STANDARD; PRT; 379 AA.
AC P55819;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Serine-glyoxylate aminotransferase (EC 2.6.1.45) (SGAT).
GN SGAA.
OS Methylobacterium extorquens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Methylobacterium group; Methylobacterium.
OX NCBI_TaxID=408;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AM1 / NCIMB 9133;
RA MEDLINE=94193542; PubMed=8144463;
RA Chistoserdova L.V., Lidstrom M.E.;
RT "Genetics of the serine cycle in Methylobacterium extorquens AM1:
RT identification of sgaA and mtdA and sequences of sgaA, hprA, and
RT mtdA.";
RL J. Bacteriol. 176:1957-1968(1994).

CC -!- CATALYTIC ACTIVITY: L-serine + glyoxylate -> 3-hydroxypyruvate +
CC glycine.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: SERINE PATHWAY.
CC -!- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L27235; -; NOT_ANNOTATED_CDS.
DR InterPro: IPR000192; Aminotransf.
DR Pfam: PF00266; aminotran_5; 1.
DR PROSITE: PS00595; AA_TRANSFER_CLASS_5; 1.
DR Transferrase: Aminotransferase; Pyridoxal phosphate.
FT BINDING 178 178 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 379 AA; 40709 MW; BEE45F4DC932A3EF CRC64;

Alignment Scores:
Pred. No.: 37.1 Length: 379
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0
DB: 1 Gaps: 0

US-09-818-990B-1 (1-3963) x SGAA_METEX (1-379)

QY 2832 AGCAATACACTTGCCCGTGGGAT 2809
|||||
DB 73 SerAsnThrLeuAlaArgGlyAsp 80

Search completed: November 30, 2002, 18:44:38
Job time : 57 secs
```


GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 30, 2002, 18:36:09 ; Search time 98.5 Seconds
(without alignments)
16579.994 Million cell updates/sec

Title: US-09-818-990B-1

Perfect score: 1316

Sequence: 1 atgcaagacagacagcataga.....tggagagtgtgaactttaa 3963

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USFTO.spool/US09818990/runat_26112002_093222_22047/app_query.fasta_1.4103
-DB=SPPREMBL_21 -OFMT=fastan -SUFFIX=oligo.rspt -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=1000 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09818990 @CGN_1.1.140 @runat_26112002_093222_22047 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rviro.*
16: sp_bacteriap.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	770	58.5	1320	4	Q96KF5

2	494	37.5	507	4	Q96K90
3	19	1.4	385	4	Q9UOF5
4	19	1.4	404	4	Q9V3E9
5	19	1.4	772	4	Q9Y2J6
6	19	1.4	1106	4	Q8WX93
7	16	1.2	379	11	Q9CWM1
8	16	1.2	492	11	Q9ET54
9	9	0.7	156	17	Q9Y6G5
10	9	0.7	239	17	Q8TXM0
11	9	0.7	243	2	Q9EV74
12	9	0.7	411	3	Q9C196
13	9	0.7	490	16	Q9K742
14	9	0.7	587	11	Q88831
15	9	0.7	663	10	Q9LT87
16	9	0.7	897	3	Q08886
17	9	0.7	1756	3	Q12193
18	9	0.7	1884	5	Q9NHW2
19	9	0.7	3469	5	Q9U412
20	9	0.7	3502	5	Q9VYJ9
21	8	0.6	48	5	Q9VWS6
22	8	0.6	102	16	Q911U1
23	8	0.6	103	12	Q99FW9
24	8	0.6	107	16	Q9RC48
25	8	0.6	110	13	Q42426
26	8	0.6	120	8	Q99132
27	8	0.6	120	17	Q8ZU19
28	8	0.6	122	16	Q9KTN9
29	8	0.6	122	17	Q58355
30	8	0.6	123	17	Q8TY74
31	8	0.6	133	10	Q9W7N3
32	8	0.6	141	11	Q61692
33	8	0.6	142	10	Q9W7N5
34	8	0.6	149	2	Q51755
35	8	0.6	149	11	Q9D4M1
36	8	0.6	153	11	Q9CZS6
37	8	0.6	154	11	Q8R331
38	8	0.6	155	4	Q96E55
39	8	0.6	164	10	Q82714
40	8	0.6	164	10	Q43489
41	8	0.6	172	6	Q95JR8
42	8	0.6	178	4	Q9H3V1
43	8	0.6	186	10	Q9FTI8
44	8	0.6	187	16	Q8XW64
45	8	0.6	192	16	Q8XGV9
46	8	0.6	192	16	Q8XD01
47	8	0.6	198	16	Q8Y5V5
48	8	0.6	198	16	Q82A58
49	8	0.6	199	16	Q8XE89
50	8	0.6	204	10	Q38895
51	8	0.6	204	16	Q9AB57
52	8	0.6	224	10	Q93244
53	8	0.6	224	11	Q9JHE6
54	8	0.6	225	4	Q9HC76
55	8	0.6	225	4	Q96FC6
56	8	0.6	225	4	Q8TEZ4
57	8	0.6	225	6	Q9BFM9
58	8	0.6	228	10	Q98S46
59	8	0.6	239	16	Q9KVL3
60	8	0.6	245	10	Q8RYE9
61	8	0.6	248	4	Q9H2W1
62	8	0.6	248	4	Q8TEZ5
63	8	0.6	248	4	Q8TBV7
64	8	0.6	249	4	Q9H2N3
65	8	0.6	253	10	Q9M5A8
66	8	0.6	255	5	Q8T857
67	8	0.6	259	10	Q94BG5
68	8	0.6	272	2	Q02304
69	8	0.6	273	16	Q8YWX4
70	8	0.6	274	16	P75019
71	8	0.6	295	16	Q8Y6Y4
72	8	0.6	304	2	Q933C0
73	8	0.6	309	12	Q8U2B4
74	8	0.6	310	2	O85727

75	8	0.6	313	10	Q9C6H3	Q9c6h3 arabidopsis	c 148	8	0.6	971	2	Q9XC25	Q9xc25 bacillus an
c 76	8	0.6	322	16	Q9WXT4	Q9wxt4 thermotoga	149	8	0.6	1066	5	Q8T3T0	Q8t3t0 dictyosteli
c 77	8	0.6	346	12	Q91FG5	Q91fg5 chilo iride	c 150	8	0.6	1070	4	Q8WX12	Q8wx12 homo sapien
c 78	8	0.6	352	10	Q04792	Q04792 garcinia ma	151	8	0.6	1095	5	Q9U4G4	Q9u4g4 drosophilla
c 79	8	0.6	356	17	Q9HS47	Q9hs47 halobacteri	c 152	8	0.6	1118	10	Q93VL6	Q93vl6 phaseolus v
c 80	8	0.6	360	2	Q48888	Q48888 mycobacteri	c 153	8	0.6	1126	5	Q8T278	Q8t278 dictyosteli
c 81	8	0.6	360	16	Q97PV7	Q97pv7 streptococ	c 154	8	0.6	1133	10	Q93VS9	Q93vs9 phaseolus v
c 82	8	0.6	362	10	Q39402	Q39402 brassica ca	155	8	0.6	1152	10	Q9C8F4	Q9c8f4 arabidopsis
c 83	8	0.6	362	10	Q42561	Q42561 arabidopsis	c 156	8	0.6	1199	12	Q83859	Q83859 nilaparvata
c 84	8	0.6	369	16	Q9PM30	Q9pm30 campylobact	157	8	0.6	1322	4	Q8TB50	Q8te50 homo sapien
c 85	8	0.6	374	2	Q47467	Q47467 erwinia car	158	8	0.6	1510	2	Q92465	Q92465 corynebacte
c 86	8	0.6	381	16	Q98QV9	Q98qv9 mycoplasma	159	8	0.6	1514	4	Q96QA9	Q96qa9 homo sapien
c 87	8	0.6	382	16	Q98CP3	Q98cp3 rhizobium l	160	8	0.6	1561	11	Q924D2	Q924d2 mus musculu
c 88	8	0.6	383	16	Q8YE66	Q8ye66 bruceella me	c 161	8	0.6	1581	16	Q92C12	Q92ci2 rickettsia
c 89	8	0.6	389	11	Q8R319	Q8r319 mus musculu	c 162	8	0.6	1583	16	Q92GE9	Q92ge9 rickettsia
c 90	8	0.6	392	13	Q98UJ7	Q98uj7 gallus gall	c 163	8	0.6	1618	11	Q9QX19	Q9qx19 rattus norv
c 91	8	0.6	415	17	Q9HMQ7	Q9hmq7 halobacteri	c 164	8	0.6	1641	11	Q88528	Q88528 mus musculu
c 92	8	0.6	416	8	Q78234	Q78234 chrysosplen	c 165	8	0.6	1754	4	Q9U1W2	Q9u1w2 homo sapien
c 93	8	0.6	424	16	Q9K044	Q9k044 neisseria m	c 166	8	0.6	1894	11	P70206	P70206 mus musculu
c 94	8	0.6	424	16	Q9JV54	Q9jv54 neisseria m	c 167	8	0.6	1905	13	Q91823	Q91823 xenopus lae
c 95	8	0.6	427	16	Q8ZFN2	Q8zfn2 versinia pe	c 168	8	0.6	2023	5	Q96542	Q96542 drosophilla
c 96	8	0.6	429	12	Q9DWH7	Q9dwh7 rat cytomeg	c 169	8	0.6	2023	5	Q9V529	Q9v529 drosophilla
c 97	8	0.6	446	13	Q13006	Q13006 ictalurus p	c 170	8	0.6	2055	11	Q88938	Q88938 mus musculu
c 98	8	0.6	456	16	Q8UDF7	Q8udf7 agrobacteri	c 171	8	0.6	2102	5	Q18183	Q18183 caenorhabdi
c 99	8	0.6	460	16	Q8UA82	Q8ua82 agrobacteri	c 172	8	0.6	2778	5	Q9V9T6	Q9v9t6 drosophilla
c 100	8	0.6	464	5	Q95SH5	Q95sh5 drosophilla	173	8	0.6	3567	11	Q9SL77	Q9es77 mus musculu
101	8	0.6	464	5	Q9W2A0	Q9w2a0 drosophilla	c 174	8	0.6	3571	10	Q9SL27	Q9sl27 arabidopsis
102	8	0.6	478	16	Q9KV20	Q9kv20 vibrio chol	c 175	8	0.6	3574	10	Q9AUB4	Q9aub4 arabidopsis
c 103	8	0.6	480	13	Q13007	Q13007 ictalurus p	c 176	8	0.6	4060	12	Q91H28	Q91h28 gill-associ
c 104	8	0.6	501	11	Q8VCW5	Q8vcw5 mus musculu	c 177	8	0.6	4924	3	Q9P7T1	Q9p7t1 schizosacch
c 105	8	0.6	508	8	Q78332	Q78332 perglularia	178	7	0.5	29	6	Q8WP11	Q8wp11 ateles belz
c 106	8	0.6	511	10	Q9XIH7	Q9xih7 arabidopsis	c 179	7	0.5	33	12	Q918E8	Q918e8 human cytom
c 107	8	0.6	511	10	Q94IH6	Q94ih6 arabidopsis	c 180	7	0.5	33	12	Q918E5	Q918e5 human cytom
c 108	8	0.6	518	10	Q94HA3	Q94ha3 oryza sativ	c 181	7	0.5	33	12	Q910Q5	Q910q5 human cytom
c 109	8	0.6	519	10	Q9STG8	Q9sti8 arabidopsis	182	7	0.5	43	2	Q46785	Q46785 escherichia
c 110	8	0.6	527	12	Q9GQ3	Q9gq3 grapevine l	183	7	0.5	43	2	Q47127	Q47127 escherichia
c 111	8	0.6	531	10	Q80735	Q80735 arabidopsis	184	7	0.5	43	2	Q57131	Q57131 escherichia
c 112	8	0.6	539	5	Q8SMW3	Q8smw3 drosophilla	185	7	0.5	45	2	Q24683	Q24683 gnetum gnet
c 113	8	0.6	559	16	Q8VRR7	Q8vrr7 anabaena sp	186	7	0.5	45	2	Q24683	Q24683 shigella bo
c 114	8	0.6	574	4	Q9H1G5	Q9h1g5 homo sapien	187	7	0.5	49	2	Q9RMR6	Q9rmr6 klebsiella
c 115	8	0.6	577	11	Q9D221	Q9dz21 mus musculu	188	7	0.5	49	16	Q8U5B4	Q8u5b4 agrobacteri
c 116	8	0.6	595	2	Q34206	Q34206 pseudomonas	c 189	7	0.5	56	7	Q9MX88	Q9mx88 oreochromis
c 117	8	0.6	595	13	Q9DBC1	Q9dbcl brachydanio	190	7	0.5	56	8	Q952Q2	Q952q2 diadema ant
c 118	8	0.6	595	16	Q9HT87	Q9ht87 pseudomonas	191	7	0.5	56	8	Q952Q0	Q952q0 diadema ant
c 119	8	0.6	598	10	Q9SSM0	Q9ssm0 arabidopsis	192	7	0.5	56	8	Q952P8	Q952p8 diadema ant
c 120	8	0.6	601	11	Q9CUT3	Q9cut3 mus musculu	193	7	0.5	56	8	Q952P7	Q952p7 diadema ant
c 121	8	0.6	628	16	Q8VHF6	Q8vhf6 bruceella me	194	7	0.5	56	8	Q952P5	Q952p5 diadema ant
c 122	8	0.6	630	3	Q13870	Q13870 schizosacch	195	7	0.5	56	8	Q952P3	Q952p3 diadema ant
c 123	8	0.6	653	15	Q87039	Q87039 simian foam	196	7	0.5	56	8	Q952P1	Q952p1 diadema ant
c 124	8	0.6	669	11	Q922H0	Q922h0 mus musculu	197	7	0.5	56	8	Q952P0	Q952p0 diadema ant
c 125	8	0.6	670	2	Q9RHQ0	Q9rhq0 pseudomonas	198	7	0.5	56	8	Q952N9	Q952n9 diadema ant
c 126	8	0.6	670	16	Q91596	Q91596 pseudomonas	199	7	0.5	56	8	Q952N7	Q952n7 diadema ant
c 127	8	0.6	674	11	Q60979	Q60379 mus musculu	200	7	0.5	56	8	Q952N5	Q952n5 diadema ant
c 128	8	0.6	680	2	Q9KHL4	Q9khl4 mycoplasma	201	7	0.5	56	8	Q952M6	Q952m6 diadema ant
c 129	8	0.6	684	16	Q917B8	Q917b8 pseudomonas	202	7	0.5	56	8	Q952M3	Q952m3 diadema ant
c 130	8	0.6	694	10	Q9LE58	Q9le58 arabidopsis	203	7	0.5	56	8	Q952J4	Q952j4 diadema ant
c 131	8	0.6	717	10	Q9LXL0	Q9lxl0 arabidopsis	204	7	0.5	56	8	Q952I5	Q952i5 diadema ant
c 132	8	0.6	717	16	Q9PBZ8	Q9pbz8 xylella fas	205	7	0.5	56	8	Q952I2	Q952i2 diadema ant
c 133	8	0.6	724	3	Q8X0Y2	Q8xy2 neurospora	206	7	0.5	56	8	Q952H6	Q952h6 diadema ant
c 134	8	0.6	748	10	Q9T0G5	Q9t0g5 arabidopsis	207	7	0.5	56	8	Q952H3	Q952h3 diadema ant
c 135	8	0.6	752	10	Q9FKZ9	Q9fkz9 arabidopsis	208	7	0.5	56	8	Q94FJ3	Q94fj3 diadema ant
c 136	8	0.6	754	5	Q8T769	Q8t769 brachiosteo	209	7	0.5	56	8	Q94N68	Q94n68 diadema ant
c 137	8	0.6	761	5	Q9VVL5	Q9vvl5 drosophilla	210	7	0.5	56	8	Q94N66	Q94n66 diadema ant
c 138	8	0.6	772	4	Q9Y468	Q9y468 homo sapien	211	7	0.5	56	8	Q94N51	Q94n51 diadema ant
c 139	8	0.6	792	5	Q9W2R3	Q9w2r3 drosophilla	212	7	0.5	56	8	Q94N50	Q94n50 diadema ant
c 140	8	0.6	793	3	Q9HFW2	Q9hfw2 ashbya goss	213	7	0.5	56	8	Q94P45	Q94p45 diadema ant
c 141	8	0.6	793	5	Q93325	Q93325 caenorhabdi	214	7	0.5	56	8	Q952N2	Q952n2 diadema ant
c 142	8	0.6	834	16	Q9PD71	Q9pd71 xylella fas	215	7	0.5	56	8	Q94N67	Q94n67 diadema ant
c 143	8	0.6	881	10	Q9SMX9	Q9smx9 arabidopsis	216	7	0.5	56	8	Q952P6	Q952p6 diadema ant
c 144	8	0.6	881	10	Q82651	Q82651 arabidopsis	217	7	0.5	56	8	Q94N49	Q94n49 diadema ant
c 145	8	0.6	887	5	Q8STF6	Q8stf6 caenorhabdi	218	7	0.5	56	8	Q8WFV4	Q8wfv4 diadema ant
c 146	8	0.6	907	16	Q9L248	Q9l248 streptomyce	219	7	0.5	56	8	Q8WF55	Q8wf55 diadema mex
c 147	8	0.6	925	2	Q9RQ00	Q9rq00 bacillus an	220	7	0.5	56	8	Q8WFO9	Q8wfv9 diadema pau

221	7	0.5	56	8	Q8WFO5	Q8wfg5 diadema pau	c 294	7	0.5	114	5	Q95Q94	Q95q94 caenorhabdi
222	7	0.5	56	8	Q8WFO4	Q8wfg4 diadema pau	c 295	7	0.5	114	16	Q99TS7	Q99ts7 staphylococ
223	7	0.5	56	8	Q8WFO2	Q8wfg2 diadema pau	c 296	7	0.5	115	12	Q902N1	Q902n1 human eryth
224	7	0.5	56	8	Q8WFO1	Q8wfg1 diadema pau	c 297	7	0.5	116	17	Q97OT8	Q97ot8 sulfolobus
225	7	0.5	56	8	Q8WFP1	Q8wfp1 diadema pau	c 298	7	0.5	117	16	Q90XP8	Q90xp8 neisseria m
226	7	0.5	56	8	Q8WFP0	Q8wfp0 diadema pau	c 299	7	0.5	117	16	Q9JW74	Q9jw74 neisseria m
227	7	0.5	56	8	Q8WFN8	Q8wfn8 diadema pau	c 300	7	0.5	117	16	Q97QR9	Q97qr9 streptococ
228	7	0.5	56	8	Q8WFN7	Q8wfn7 diadema sav	c 301	7	0.5	118	2	Q9AQG0	Q9aqg0 acidovorax
229	7	0.5	56	8	Q8WFM2	Q8wfm2 diadema sav	c 302	7	0.5	118	4	Q9NOF6	Q9nof6 homo sapien
230	7	0.5	56	8	Q8WFM4	Q8wfm4 diadema ant	c 303	7	0.5	118	5	Q95X43	Q95x43 caenorhabdi
231	7	0.5	56	8	Q8W7R7	Q8w7r7 diadema sav	c 304	7	0.5	118	5	Q9UIK5	Q9uik5 drosophila
232	7	0.5	56	8	Q8W7L8	Q8w7l8 diadema mex	c 305	7	0.5	118	8	Q9ME35	Q9me35 oxylooma hay
233	7	0.5	56	8	Q8W7L6	Q8w7l6 diadema sav	c 306	7	0.5	118	8	Q9MEP0	Q9mep0 oxylooma hay
234	7	0.5	56	8	Q8W7C9	Q8w7c9 diadema pau	c 307	7	0.5	118	10	Q49947	Q49947 solanum tub
235	7	0.5	56	8	Q8W7C8	Q8w7c8 diadema sav	c 308	7	0.5	118	16	Q8R8Y7	Q8r8y7 thermoaer
236	7	0.5	65	6	Q97832	Q97832 canis famil	c 309	7	0.5	119	5	Q9V401	Q9v401 drosophila
237	7	0.5	65	7	Q77940	Q77940 sus scrofa	c 310	7	0.5	119	5	Q9V8F3	Q9v8f3 drosophila
238	7	0.5	68	10	Q9S967	Q9s967 medicago sa	c 311	7	0.5	119	10	Q9ZSF5	Q9zsf5 lycium ande
239	7	0.5	69	2	Q57052	Q57052 staphylococ	c 312	7	0.5	120	4	Q75591	Q75591 homo sapien
240	7	0.5	70	2	Q9EXE4	Q9exe4 lactobacilli	c 313	7	0.5	120	4	Q13542	Q13542 homo sapien
241	7	0.5	70	10	Q8S863	Q8s863 oryza sativ	c 314	7	0.5	120	16	Q8Y379	Q8y379 ralstonia s
242	7	0.5	71	10	Q8S8P7	Q8s8p7 arabidopsis	c 315	7	0.5	121	12	Q9YRJ5	Q9yrl5 human cytom
243	7	0.5	71	16	Q9JQW9	Q9jqw9 neisseria m	c 316	7	0.5	121	17	Q97VK2	Q97vk2 sulfolobus
244	7	0.5	73	8	Q20453	Q20453 fugu rubrip	c 317	7	0.5	122	12	Q9YRJ1	Q9yrl1 human cytom
245	7	0.5	73	8	Q9XQR3	Q9xqr3 pisum sativ	c 318	7	0.5	122	16	Q8YAW3	Q8yaw3 bruceella me
246	7	0.5	75	16	Q92NY7	Q92ny7 rhizobium m	c 319	7	0.5	123	2	Q32992	Q32992 mycobacteri
247	7	0.5	79	12	Q68100	Q68100 human cytom	c 320	7	0.5	123	8	Q9MJM2	Q9mjm2 limnodynast
248	7	0.5	79	16	Q8XXZ3	Q8xxz3 ralstonia s	c 321	7	0.5	123	11	Q9CZ38	Q9cz38 mus musculu
249	7	0.5	80	10	Q9CA50	Q9ca50 arabidopsis	c 322	7	0.5	123	12	Q9DJU9	Q9dju9 cherry gree
250	7	0.5	80	16	Q9PP50	Q9pp50 campylobact	c 323	7	0.5	124	5	Q95RN4	Q95rn4 drosophila
251	7	0.5	82	15	Q9YKY7	Q9yky7 human immun	c 324	7	0.5	124	10	Q92PV4	Q92pv4 arabidopsis
252	7	0.5	83	2	Q9RDW4	Q9rdw4 lactobacilli	c 325	7	0.5	124	17	Q976C8	Q976c8 sulfolobus
253	7	0.5	83	2	Q8VRG2	Q8vrg2 lactobacilli	c 326	7	0.5	124	17	Q26783	Q26783 methanobact
254	7	0.5	85	11	Q8S095	Q8s095 mus musculu	c 327	7	0.5	125	6	Q95K14	Q95k14 macaca fasc
255	7	0.5	87	5	Q94787	Q94787 trypanosoma	c 328	7	0.5	125	8	Q9XNM4	Q9xnm4 tupinambis
256	7	0.5	88	17	Q970B1	Q970b1 sulfolobus	c 329	7	0.5	125	8	Q9TFC1	Q9tfc1 tupinambis
257	7	0.5	89	5	Q9VS42	Q9vs42 drosophila	c 330	7	0.5	125	8	Q9TFE9	Q9tfe9 tupinambis
258	7	0.5	90	8	Q953X0	Q953x0 terebratali	c 331	7	0.5	125	8	Q9XN43	Q9xn43 cnemidophor
259	7	0.5	92	16	Q9JXN2	Q9jxn2 neisseria m	c 332	7	0.5	125	8	Q9XN42	Q9xn42 ameiva amei
260	7	0.5	95	10	Q8S2N1	Q8s2n1 oryza sativ	c 333	7	0.5	125	8	Q9XN41	Q9xn41 kentropyx v
261	7	0.5	95	16	Q9K102	Q9k102 neisseria m	c 334	7	0.5	125	8	Q9XN40	Q9xn40 tupinambis
262	7	0.5	96	16	Q8RHS9	Q8rhs9 fusobacteri	c 335	7	0.5	125	8	Q9TFB9	Q9tfb9 tupinambis
263	7	0.5	98	12	Q9YR19	Q9yrl9 human cytom	c 336	7	0.5	125	8	Q9XK25	Q9xk25 tupinambis
264	7	0.5	100	3	Q98681	Q98681 saccharomyc	c 337	7	0.5	125	8	Q9XJZ9	Q9xjz9 tupinambis
265	7	0.5	100	12	Q9YRJ4	Q9yrl4 human cytom	c 338	7	0.5	125	8	Q9XJZ8	Q9xjz8 tupinambis
266	7	0.5	100	16	Q9JW91	Q9jw91 neisseria m	c 339	7	0.5	125	8	Q9T3U5	Q9t3u5 tupinambis
267	7	0.5	100	17	Q9TDP9	Q9tdp9 aeropyrum p	c 340	7	0.5	125	8	Q9T3M0	Q9t3m0 tupinambis
268	7	0.5	101	12	Q9YR17	Q9yrl7 human cytom	c 341	7	0.5	125	17	Q9YCW4	Q9ycw4 aeropyrum p
269	7	0.5	101	15	Q9YKR3	Q9ykr3 human immun	c 342	7	0.5	126	16	Q8YKM8	Q8ykm8 anabaena sp
270	7	0.5	102	2	Q50225	Q50225 thiobacillu	c 343	7	0.5	126	16	Q8YKD6	Q8ykd6 anabaena sp
271	7	0.5	102	16	Q9RTH8	Q9rth8 deinococcus	c 344	7	0.5	128	2	Q8RL04	Q8rl04 providencia
272	7	0.5	103	15	Q8UT31	Q8ut31 human immun	c 345	7	0.5	128	3	Q9HGT1	Q9hgt1 coccidioidi
273	7	0.5	104	2	Q45350	Q45350 beta proteo	c 346	7	0.5	128	5	Q9VME4	Q9vme4 drosophila
274	7	0.5	104	5	Q8SVF6	Q8svf6 encephalito	c 347	7	0.5	128	8	Q9MJN8	Q9mjn8 limnodynast
275	7	0.5	105	5	Q9VPE5	Q9vpe5 drosophila	c 348	7	0.5	128	12	Q68084	Q68084 human cytom
276	7	0.5	105	5	Q8T3M1	Q8t3m1 drosophila	c 349	7	0.5	129	2	Q52356	Q52356 mycoplasma
277	7	0.5	105	11	Q9D4U9	Q9d4u9 mus musculu	c 350	7	0.5	129	4	Q9BXM8	Q9bxm8 homo sapien
278	7	0.5	105	11	Q9S774	Q9s774 mus musculu	c 351	7	0.5	130	4	Q9BSB0	Q9bsb0 homo sapien
279	7	0.5	105	12	Q9YRJ6	Q9yrl6 human cytom	c 352	7	0.5	130	15	Q76870	Q76870 human immun
280	7	0.5	106	12	Q9WKR3	Q9wkr3 human cytom	c 353	7	0.5	130	15	Q76871	Q76871 human immun
281	7	0.5	106	12	Q88468	Q88468 strawberry	c 354	7	0.5	130	15	Q76872	Q76872 human immun
282	7	0.5	107	16	Q9JXU1	Q9jxu1 neisseria m	c 355	7	0.5	130	15	Q76873	Q76873 human immun
283	7	0.5	109	6	Q28118	Q28118 bos taurus	c 356	7	0.5	130	15	Q76874	Q76874 human immun
284	7	0.5	110	10	Q94HZ7	Q94hz7 oryza sativ	c 357	7	0.5	130	15	Q76877	Q76877 human immun
285	7	0.5	110	17	Q58195	Q58195 pyrococcus	c 358	7	0.5	130	15	Q76878	Q76878 human immun
286	7	0.5	111	5	Q96509	Q96509 caenorhabdi	c 359	7	0.5	130	15	Q76879	Q76879 human immun
287	7	0.5	111	11	Q88435	Q88435 rattus norv	c 360	7	0.5	130	15	Q76880	Q76880 human immun
288	7	0.5	111	12	Q9YRJ8	Q9yrl8 human cytom	c 361	7	0.5	130	15	Q76881	Q76881 human immun
289	7	0.5	111	16	Q9RTG6	Q9rtg6 deinococcus	c 362	7	0.5	130	15	Q76900	Q76900 human immun
290	7	0.5	112	11	Q8VDH3	Q8vdh3 mus musculu	c 363	7	0.5	130	15	Q76901	Q76901 human immun
291	7	0.5	112	12	Q9YRJ7	Q9yrl7 human cytom	c 364	7	0.5	130	15	Q76902	Q76902 human immun
292	7	0.5	113	16	Q98MW1	Q98mw1 rhizobium l	c 365	7	0.5	130	15	Q76903	Q76903 human immun
293	7	0.5	113	16	Q98K14	Q98k14 rhizobium l	c 366	7	0.5	130	15	Q76904	Q76904 human immun

c 367	7	0.5	130	15	Q76905	Q76905 human immun	c 440	7	0.5	145	13	Q9DG79	Q9dg79 ictalurus p
c 368	7	0.5	130	15	Q76907	Q76907 human immun	c 441	7	0.5	146	15	Q90BC4	Q90bc4 human immun
c 369	7	0.5	130	15	Q76908	Q76908 human immun	442	7	0.5	146	15	P94515	P94515 bacillus su
c 370	7	0.5	130	15	Q76909	Q76909 human immun	443	7	0.5	146	16	Q8UBR4	Q8ubr4 agrobacteri
c 371	7	0.5	130	15	Q76910	Q76910 human immun	444	7	0.5	146	16	Q8R708	Q8r708 thermoanaer
c 372	7	0.5	130	15	Q77049	Q77049 human immun	c 445	7	0.5	147	5	Q9GYM0	Q9gym0 caenorhabdi
c 373	7	0.5	130	15	Q77050	Q77050 human immun	c 446	7	0.5	148	4	Q9UWJ4	Q9umj4 homo sapien
c 374	7	0.5	130	15	Q77052	Q77052 human immun	c 447	7	0.5	148	10	Q49179	Q49179 oryza sativ
c 375	7	0.5	130	15	Q77055	Q77055 human immun	448	7	0.5	148	10	Q8VXY1	Q8vxy1 arabidopsis
c 376	7	0.5	130	15	Q77057	Q77057 human immun	449	7	0.5	148	10	Q8VXY1	Q8vxy1 arabidopsis
c 377	7	0.5	130	15	Q77058	Q77058 human immun	c 450	7	0.5	149	6	Q9TVB4	Q9tvb4 canis famill
c 378	7	0.5	130	15	Q77088	Q77088 human immun	451	7	0.5	150	8	Q9MLH3	Q9mlh3 paragonimus
c 379	7	0.5	130	15	Q77089	Q77089 human immun	c 452	7	0.5	150	8	Q9MLH3	Q9mlh3 paragonimus
380	7	0.5	131	8	Q9MJM4	Q9mjM4 limnodynast	c 453	7	0.5	150	16	Q55491	Q55491 synochocyst
381	7	0.5	131	10	Q9XES9	Q9xes9 glycine max	454	7	0.5	150	16	Q9G011	Q9g011 rhizobium l
382	7	0.5	131	10	Q9LKU5	Q9lku5 arabidopsis	c 455	7	0.5	150	16	Q921J6	Q921j6 rickettsia
383	7	0.5	132	8	Q9MJN3	Q9mjN3 limnodynast	456	7	0.5	151	2	Q68739	Q68739 yersinia pe
384	7	0.5	134	9	Q34035	Q34035 streptococc	457	7	0.5	151	10	Q9FWES	Q9fwe5 oryza sativ
c 385	7	0.5	134	10	Q94GZ2	Q94gz2 oryza sativ	c 458	7	0.5	151	16	Q84177	Q84177 chlamydia t
c 386	7	0.5	134	16	Q9RSP5	Q9rsp5 deinococcus	c 459	7	0.5	151	16	Q9CGK9	Q9cgk9 lactococcus
c 387	7	0.5	135	5	Q17248	Q17248 caenorhabdi	460	7	0.5	152	5	Q95ZL9	Q95z19 caenorhabdi
388	7	0.5	135	8	Q9MJP7	Q9mjp7 crinia dese	c 461	7	0.5	152	16	Q9ABX2	Q9abx2 caulobacter
389	7	0.5	135	8	Q9MJP6	Q9mjp6 crinia pari	462	7	0.5	153	5	Q9UIK3	Q9ulK3 drosophila
c 390	7	0.5	135	10	Q9FXQ7	Q9fxq7 zea mays (m	463	7	0.5	153	5	Q9VFL5	Q9vfl5 drosophila
c 391	7	0.5	135	16	Q9RJZ5	Q9rjz5 streptomyce	c 464	7	0.5	154	5	Q8T972	Q8t972 drosophila
392	7	0.5	136	8	Q9ME19	Q9me19 limnodynast	465	7	0.5	154	16	Q8UA14	Q8ua14 agrobacteri
393	7	0.5	136	8	Q9MDW5	Q9mdw5 limnodynast	c 466	7	0.5	154	16	Q9Y8V9	Q9y8v9 aeropyrum p
394	7	0.5	136	8	Q9MDU2	Q9mdu2 limnodynast	c 467	7	0.5	155	5	Q9TXS7	Q9txs7 caenorhabdi
395	7	0.5	136	8	Q9MDS3	Q9mds3 limnodynast	c 468	7	0.5	156	10	Q41718	Q41718 zea diplope
396	7	0.5	136	8	Q9MDS2	Q9mds2 limnodynast	c 469	7	0.5	156	10	Q93YF6	Q93yf6 nicotiana s
397	7	0.5	136	8	Q9MDM3	Q9mdm3 limnodynast	470	7	0.5	156	15	Q92813	Q92813 bovine leuk
398	7	0.5	136	8	Q9MJN9	Q9mjN9 limnodynast	c 471	7	0.5	156	16	Q9A9N5	Q9a9n5 caulobacter
399	7	0.5	136	8	Q9MJP5	Q9mjp5 adelotus br	472	7	0.5	157	5	Q9WLK2	Q9wlk2 drosophila
400	7	0.5	136	8	Q9MJP4	Q9mjp4 megistoloti	c 473	7	0.5	157	16	Q9PIP3	Q9pip3 campylobact
401	7	0.5	136	8	Q9MJP1	Q9mjp1 limnodynast	474	7	0.5	157	16	Q9P9P4	Q9ppp4 xylella fas
402	7	0.5	136	8	Q9MJN7	Q9mjN7 limnodynast	475	7	0.5	158	10	Q941E5	Q941e5 arabidopsis
403	7	0.5	136	8	Q9MJN6	Q9mjN6 limnodynast	476	7	0.5	158	12	Q56249	Q56249 tobacco rat
404	7	0.5	136	8	Q9MJN5	Q9mjN5 limnodynast	c 477	7	0.5	159	10	Q22690	Q22690 arabidopsis
405	7	0.5	136	8	Q9MJN4	Q9mjN4 limnodynast	c 478	7	0.5	159	17	Q97YG6	Q97yg6 sulfolobus
406	7	0.5	136	8	Q9MJN2	Q9mjN2 limnodynast	c 479	7	0.5	160	8	Q8WAG1	Q8wag1 sphaerospir
407	7	0.5	136	8	Q9MJN1	Q9mjN1 limnodynast	c 480	7	0.5	160	8	Q8WAG0	Q8wag0 sphaerospir
408	7	0.5	136	8	Q9MJN0	Q9mjN0 limnodynast	c 481	7	0.5	160	8	Q8WAF9	Q8waf9 sphaerospir
409	7	0.5	136	8	Q9MJM9	Q9mjM9 limnodynast	c 482	7	0.5	160	8	Q8WAF8	Q8waf8 sphaerospir
410	7	0.5	136	8	Q9MJM8	Q9mjM8 limnodynast	c 483	7	0.5	160	16	Q9KM07	Q9km07 vibrio chol
411	7	0.5	136	8	Q9MJM7	Q9mjM7 limnodynast	c 484	7	0.5	160	17	Q58511	Q58511 pyrococcus
412	7	0.5	136	8	Q9MJM6	Q9mjM6 limnodynast	485	7	0.5	160	17	Q97WV2	Q97wv2 sulfolobus
413	7	0.5	136	8	Q9MJM5	Q9mjM5 limnodynast	c 486	7	0.5	161	4	Q9NWZ0	Q9nwz0 homo sapien
414	7	0.5	136	8	Q9MJM3	Q9mjM3 limnodynast	c 487	7	0.5	161	4	Q9NT07	Q9nty07 homo sapien
415	7	0.5	136	8	Q9MJM1	Q9mjM1 limnodynast	c 488	7	0.5	161	4	Q9BVLL	Q9bvll homo sapien
c 416	7	0.5	136	16	Q9JDM3	Q9jdm3 neisseria m	489	7	0.5	161	16	Q9PCH9	Q9pch9 xylella fas
c 417	7	0.5	137	12	Q9E2L6	Q9e2l6 hepatitis c	490	7	0.5	161	17	Q26735	Q26735 methanobact
418	7	0.5	137	12	Q919L8	Q919l8 culex nigri	c 491	7	0.5	162	5	Q23137	Q23137 caenorhabdi
419	7	0.5	137	16	Q8ZFB8	Q8zfb8 yersinia pe	c 492	7	0.5	162	8	Q95AX6	Q95ax6 cyclanthus
c 420	7	0.5	138	11	Q9CUP0	Q9cup0 mus musculus	c 493	7	0.5	162	8	Q95AX0	Q95ax0 pandanus ut
421	7	0.5	138	12	Q9YR15	Q9yr15 human cytom	494	7	0.5	162	16	Q8XQU9	Q8xqu9 ralstonia s
c 422	7	0.5	138	15	Q41391	Q41391 human immun	c 495	7	0.5	163	17	Q9YGL1	Q9ygl1 aeropyrum p
423	7	0.5	138	16	Q9RX00	Q9rx00 deinococcus	c 496	7	0.5	164	2	Q9EUJ9	Q9euj9 salmonella
c 424	7	0.5	138	16	Q9A9G5	Q9a9g5 caulobacter	c 497	7	0.5	164	2	Q70017	Q70017 shigella dy
425	7	0.5	138	16	Q8UG10	Q8ug10 agrobacteri	c 498	7	0.5	164	4	Q96QK6	Q96qk6 homo sapien
426	7	0.5	139	10	Q9LNI6	Q9lni6 arabidopsis	499	7	0.5	164	10	Q8WZ24	Q8wz24 oryza sativ
427	7	0.5	139	12	Q9YRJ2	Q9yrj2 human cytom	c 500	7	0.5	164	16	Q8X5N5	Q8x5n5 escherichia
c 428	7	0.5	139	13	Q9DG76	Q9dg76 ictalurus p	c 501	7	0.5	164	16	Q8VKH4	Q8vkh4 mycobacteri
429	7	0.5	139	16	Q8Z5Q2	Q8z5q2 salmonella	502	7	0.5	165	15	Q91G02	Q91g02 bovine leuk
430	7	0.5	140	10	Q8SJH1	Q8sjh1 oryza sativ	c 503	7	0.5	165	16	Q97IH9	Q97ih9 clostridium
431	7	0.5	140	12	Q9QU35	Q9qu35 ttv-like ml	c 504	7	0.5	165	16	Q9X9V2	Q9x9v2 streptomyce
c 432	7	0.5	142	6	Q9GKN3	Q9gkn3 bos taurus	505	7	0.5	166	16	Q9A3E6	Q9a3e6 caulobacter
433	7	0.5	143	5	Q9N7Y0	Q9n7y0 leishmania	506	7	0.5	166	16	Q9ADM6	Q9adm6 streptomyce
434	7	0.5	143	12	Q90189	Q90189 bombyx mori	c 507	7	0.5	168	2	Q93IK3	Q93ik3 vibrio sp.
c 435	7	0.5	143	16	Q8YHU9	Q8yhu9 brucella me	c 508	7	0.5	168	16	Q92S15	Q92s15 rhizobium m
436	7	0.5	144	5	Q9GZ13	Q9gz13 caenorhabdi	c 509	7	0.5	169	8	Q94S76	Q94s76 zenopsis ne
c 437	7	0.5	145	2	Q9AP12	Q9ap12 legionella	c 510	7	0.5	169	12	Q912C6	Q912c6 human eryth
c 438	7	0.5	145	5	Q9VTF3	Q9vtf3 drosophila	c 511	7	0.5	170	2	Q9KKI9	Q9kki9 yersinia en
c 439	7	0.5	145	11	Q9D6U0	Q9d6u0 mus musculus	512	7	0.5	170	16	Q8R6X4	Q8r6x4 thermoanaer

513	7	0.5	171	4	O00486	O00486 homo sapien	586	7	0.5	185	8	Q94Q98	Q94Q98 limnodynast
514	7	0.5	171	16	O8UIR4	O8uir4 agrobacteri	587	7	0.5	185	8	Q94Q64	Q94Q64 limnodynast
515	7	0.5	172	5	O61948	O61948 caenorhabdi	588	7	0.5	185	8	Q94PI0	Q94pi0 limnodynast
516	7	0.5	172	8	Q966Q2	Q966q2 diplophos t	589	7	0.5	185	8	Q94NN7	Q94nn7 limnodynast
517	7	0.5	173	8	Q966N9	Q966n9 polymixia j	590	7	0.5	185	8	Q94NM0	Q94nm0 limnodynast
518	7	0.5	173	8	Q96B9U4	Q96b9u4 aulopus jep	591	7	0.5	185	8	Q94NG7	Q94ng7 limnodynast
519	7	0.5	173	8	Q94YR3	Q94yr3 engraulis j	592	7	0.5	185	8	Q94NG6	Q94ng6 limnodynast
520	7	0.5	173	8	Q94T69	Q94t69 lampris gut	593	7	0.5	185	8	Q94N10	Q94n10 limnodynast
521	7	0.5	173	8	Q94T35	Q94t35 polymixia l	594	7	0.5	185	16	Q9CHS8	Q9chs8 lactococcus
522	7	0.5	173	8	Q94SK9	Q94sk9 zeus faber	595	7	0.5	186	4	Q96ST6	Q96st6 homo sapien
523	7	0.5	173	8	Q94SF7	Q94sf7 monopteris	c 596	7	0.5	186	8	Q9MIX5	Q9mix5 uroleucon s
524	7	0.5	174	2	Q9EYX7	Q9eyx7 klebsiella	597	7	0.5	186	8	Q956V1	Q956v1 limnodynast
525	7	0.5	174	2	Q8RTW3	Q8rtw3 uncultured	598	7	0.5	186	8	Q956U4	Q956u4 limnodynast
526	7	0.5	174	4	Q9NSU8	Q9nsu8 homo sapien	599	7	0.5	186	10	Q94IM0	Q94im0 oryza sativ
527	7	0.5	174	5	Q9NGY3	Q9ngy3 difofilaria	600	7	0.5	186	16	Q9PB72	Q9pb72 chlamydia p
528	7	0.5	174	5	O76968	O76968 podocoryne	601	7	0.5	186	16	Q8RHL3	Q8rhl3 fusobacteri
529	7	0.5	174	11	Q9CTX3	Q9ctx3 mus musculu	602	7	0.5	186	17	O27711	O27711 methanobact
530	7	0.5	175	5	Q23646	Q23646 caenorhabdi	603	7	0.5	187	10	O24352	O24352 silene lati
531	7	0.5	175	11	Q9D6Z3	Q9d6z3 mus musculu	604	7	0.5	187	10	O24020	O24020 lycopersico
532	7	0.5	175	11	Q9D6W5	Q9d6w5 mus musculu	605	7	0.5	188	5	Q9W012	Q9w012 drosophila
533	7	0.5	175	17	O8TP57	O8tp57 methanosarc	606	7	0.5	188	6	O8WMD0	O8wmd0 canis famil
534	7	0.5	176	6	P79329	P79329 callitrix	607	7	0.5	188	10	Q9MAV1	Q9mav1 arabidopsis
535	7	0.5	176	16	Q9ABZ9	Q9abz9 caulobacter	608	7	0.5	188	16	Q9A3H5	Q9a3h5 caulobacter
536	7	0.5	176	16	Q8RI01	Q8ri01 fusobacteri	609	7	0.5	189	2	Q54918	Q54918 streptococc
537	7	0.5	177	4	Q96IM9	Q96im9 homo sapien	610	7	0.5	189	4	Q96GS0	Q96gs0 homo sapien
538	7	0.5	177	8	Q956T0	Q956t0 limnodynast	611	7	0.5	189	16	Q9B5I8	Q9b5i8 rhizobium l
539	7	0.5	177	12	O8QLJ4	O8qlj4 mamestra co	612	7	0.5	189	16	O8Z614	O8z614 salmonella
540	7	0.5	177	17	Q9YCX1	Q9ycx1 aeropyrum p	613	7	0.5	189	16	O31771	O31771 bacillus su
541	7	0.5	178	2	Q54933	Q54933 streptococc	614	7	0.5	190	4	O8TPP5	O8tp55 mus sapien
542	7	0.5	178	8	Q956U9	Q956u9 limnodynast	615	7	0.5	190	5	Q9V5P2	Q9v5p2 drosophila
543	7	0.5	178	10	O48681	O48681 arabidopsis	616	7	0.5	190	6	O8WNM1	O8wnm1 gorilla gor
544	7	0.5	178	10	O8RVQ9	O8rvq9 arabidopsis	617	7	0.5	190	6	O8WNM0	O8wnm0 pongo pygma
545	7	0.5	178	11	O9DA50	O9da50 mus musculu	618	7	0.5	190	8	O9B3V2	O9b3v2 plethodon g
546	7	0.5	178	16	Q9RT80	Q9rt80 deinococcus	619	7	0.5	190	8	O9B3U4	O9b3u4 plethodon t
547	7	0.5	178	16	Q9HT17	Q9ht17 pseudomonas	620	7	0.5	190	8	O8WC80	O8wc80 anotosaura
548	7	0.5	179	10	Q9C571	Q9c571 arabidopsis	621	7	0.5	190	11	Q9D3L0	Q9d3l0 mus musculu
549	7	0.5	179	16	Q99SF2	Q99sf2 staphylococ	622	7	0.5	190	11	Q9D238	Q9d238 mus musculu
550	7	0.5	180	5	Q99968	Q99968 caenorhabdi	623	7	0.5	190	11	O9D052	Q9d052 mus musculu
551	7	0.5	180	5	O18997	O18997 caenorhabdi	624	7	0.5	190	11	O9CYF8	Q9cyf8 mus musculu
552	7	0.5	180	8	Q956U2	Q956u2 limnodynast	625	7	0.5	190	11	O8VDR1	O8vdr1 mus musculu
553	7	0.5	180	16	Q9KMQ7	Q9kmq7 vibrio chol	626	7	0.5	190	12	Q90149	Q90149 bombyx mori
554	7	0.5	180	16	Q9CFU5	Q9cfu5 lactococcus	627	7	0.5	191	2	O51793	O51793 ralstonia m
555	7	0.5	180	16	O8G314	O8g314 mycobacteri	628	7	0.5	191	4	Q96MY2	Q96my2 homo sapien
556	7	0.5	180	16	O8UJ07	O8uj07 agrobacteri	629	7	0.5	192	8	O9B3U8	O9b3u8 plethodon p
557	7	0.5	180	16	O8UID9	O8uid9 agrobacteri	630	7	0.5	192	8	O8WC81	O8wc81 alopoglossu
558	7	0.5	181	5	O9GZ14	O9gz14 caenorhabdi	631	7	0.5	192	11	O8R3W8	O8r3w8 mus musculu
559	7	0.5	181	8	Q956T2	Q956t2 limnodynast	632	7	0.5	192	16	Q9I2L9	Q9i2l9 pseudomonas
560	7	0.5	181	10	Q9FGF1	Q9fgf1 arabidopsis	633	7	0.5	193	16	Q9JWK5	Q9jwk5 neisseria m
561	7	0.5	181	10	Q9LYH4	Q9lyh4 arabidopsis	634	7	0.5	194	5	Q9V827	Q9v827 drosophila
562	7	0.5	182	5	O20866	O20866 caenorhabdi	635	7	0.5	194	8	O8WC83	O8wc83 placosoma g
563	7	0.5	182	5	O9VCG7	O9vcg7 drosophila	636	7	0.5	194	8	O8WC82	O8wc82 alopoglossu
564	7	0.5	182	8	Q956U7	Q956u7 limnodynast	637	7	0.5	194	10	O8VXC4	O8vxc4 oryza sativ
565	7	0.5	182	8	Q956T9	Q956t9 limnodynast	638	7	0.5	195	4	O9H7R4	O9h7r4 homo sapien
566	7	0.5	182	8	Q956T1	Q956t1 limnodynast	639	7	0.5	195	5	Q9V9Z1	Q9v9z1 drosophila
567	7	0.5	182	10	Q93ZG8	Q93zg8 arabidopsis	640	7	0.5	195	16	Q9RWJ5	Q9rwj5 deinococcus
568	7	0.5	182	11	Q9ER90	Q9er90 mus musculu	641	7	0.5	196	11	O8R1H7	O8r1h7 mus musculu
569	7	0.5	182	16	O9I539	O9i539 pseudomonas	642	7	0.5	197	8	O8WC84	O8wc84 gymnoptthal
570	7	0.5	183	8	Q956U6	Q956u6 limnodynast	643	7	0.5	197	10	Q9S9P9	Q9s9p9 arabidopsis
571	7	0.5	184	8	Q956T7	Q956t7 limnodynast	644	7	0.5	197	16	O06983	O06983 bacillus su
572	7	0.5	185	8	Q956V0	Q956v0 limnodynast	645	7	0.5	197	16	O9JZK3	O9jzk3 neisseria m
573	7	0.5	185	8	Q956U8	Q956u8 limnodynast	650	7	0.5	197	16	O92K09	O92ku9 rhizobium m
574	7	0.5	185	8	Q956U5	Q956u5 limnodynast	646	7	0.5	197	16	O92K09	O92ku9 rhizobium m
575	7	0.5	185	8	Q956U3	Q956u3 limnodynast	647	7	0.5	198	2	Q9F419	Q9f419 mycobacteri
576	7	0.5	185	8	Q956U1	Q956u1 limnodynast	648	7	0.5	198	4	Q966Y7	Q966y7 homo sapien
577	7	0.5	185	8	Q956U0	Q956u0 limnodynast	649	7	0.5	198	5	O45186	O45186 caenorhabdi
578	7	0.5	185	8	Q956T8	Q956t8 limnodynast	650	7	0.5	198	8	O8WCC5	O8wcc5 alopoglossu
579	7	0.5	185	8	Q956T6	Q956t6 limnodynast	651	7	0.5	198	10	Q9M844	Q9m844 arabidopsis
580	7	0.5	185	8	Q956T5	Q956t5 limnodynast	652	7	0.5	199	2	Q9RWF1	Q9rmf1 zymomonas m
581	7	0.5	185	8	Q956T4	Q956t4 limnodynast	653	7	0.5	200	4	O00485	O00485 homo sapien
582	7	0.5	185	8	Q956T3	Q956t3 limnodynast	654	7	0.5	200	8	O9B3U3	O9b3u3 plethodon y
583	7	0.5	185	8	Q956S9	Q956s9 limnodynast	655	7	0.5	200	8	O9B3U1	O9b3u1 plethodon w
584	7	0.5	185	8	Q956S8	Q956s8 limnodynast	656	7	0.5	200	8	O959Q7	O959q7 dendrodoris
585	7	0.5	185	8	Q94Q99	Q94q99 limnodynast	657	7	0.5	200	8	O959Q3	O959q3 dendrodoris
							658	7	0.5	200	16	Q99TE4	Q99te4 staphylococ

659	7	0.5	200	16	Q8YBB1	Q8Ybb1 brucella me	c 732	7	0.5	211	16	Q98410	Q98410 rhizobium l
c 660	7	0.5	201	10	P93539	P93539 silene prat	733	7	0.5	211	17	Q9YCC2	Q9Ycc2 aeropyrum p
c 661	7	0.5	201	16	Q9AD42	Q9ad42 streptomyce	c 734	7	0.5	211	17	Q8TLT7	Q8Tlt7 methanosarc
662	7	0.5	202	8	Q8WC85	Q8wc85 neusticurus	735	7	0.5	212	8	Q9B0K3	Q9B0k3 limnodynast
c 663	7	0.5	202	11	Q9DAP7	Q9dap7 mus musculus	c 736	7	0.5	212	12	Q41980	Q41980 murid herpe
c 664	7	0.5	202	16	Q9A841	Q9a841 caulobacter	737	7	0.5	214	5	Q8WSH4	Q8Wsh4 ancylostoma
665	7	0.5	203	8	Q9B3V3	Q9b3v3 plethodon f	738	7	0.5	214	8	Q9GCB6	Q9Gcb6 litoria cae
666	7	0.5	203	8	Q8WC88	Q8wc88 anotosaura	739	7	0.5	214	8	Q9GCB5	Q9Gcb5 litoria cit
667	7	0.5	203	8	Q8WC87	Q8wc87 calyptommat	740	7	0.5	214	8	Q9GCB4	Q9Gcb4 litoria cit
668	7	0.5	203	8	Q8WC86	Q8wc86 micrablepha	741	7	0.5	214	8	Q9GCB3	Q9Gcb3 litoria sub
c 669	7	0.5	203	11	Q9CV13	Q9cv13 mus musculus	742	7	0.5	214	8	Q9GCB2	Q9Gcb2 litoria sub
670	7	0.5	204	2	Q93IH4	Q93ih4 wolinnella s	743	7	0.5	214	8	Q9GCB1	Q9Gcb1 litoria sp.
c 671	7	0.5	204	10	Q23660	Q23660 arabidopsis	744	7	0.5	214	8	Q9GCB0	Q9Gcb0 litoria spe
c 672	7	0.5	204	11	Q9QX18	Q9qx18 rattus norv	745	7	0.5	214	8	Q9GCA9	Q9Gca9 litoria nud
c 673	7	0.5	204	12	Q66087	Q66087 canine herp	746	7	0.5	214	8	Q9GCA8	Q9Gca8 litoria phy
c 674	7	0.5	204	16	Q8YGU3	Q8ygu3 brucella me	747	7	0.5	214	8	Q9GCA7	Q9Gca7 litoria pea
c 675	7	0.5	205	2	Q51712	Q51712 paracoccus	748	7	0.5	214	8	Q9GCA6	Q9Gca6 litoria pea
676	7	0.5	205	8	Q9B3V0	Q9b3v0 plethodon j	749	7	0.5	214	8	Q9B0J7	Q9B0j7 limnodynast
677	7	0.5	205	8	Q8WC89	Q8wc89 eclepousus g	750	7	0.5	214	8	Q952E2	Q952e2 lineatriton
678	7	0.5	206	5	Q9NM09	Q9nm09 leishmania	c 751	7	0.5	214	10	Q94H35	Q94h35 oryza sativ
679	7	0.5	206	5	Q9VYC9	Q9vyc9 drosophila	752	7	0.5	214	10	Q9SV19	Q9sv19 arabidopsis
680	7	0.5	206	8	Q8WC92	Q8wc92 leposoma pe	c 753	7	0.5	215	2	Q93A10	Q93a10 thiobacilli
681	7	0.5	207	8	Q9B019	Q9b019 limnodynast	c 754	7	0.5	216	3	Q9USA7	Q9usa7 schizosacch
682	7	0.5	207	8	Q8WCC4	Q8wcc4 arthrosaura	755	7	0.5	216	8	Q9B0L0	Q9B0l0 limnodynast
683	7	0.5	207	8	Q8WCC3	Q8wcc3 gymnoththal	756	7	0.5	216	16	Q98H77	Q98h77 rhizobium l
684	7	0.5	207	8	Q8WCC2	Q8wcc2 colobosaura	757	7	0.5	217	3	Q94290	Q94290 schizosacch
685	7	0.5	207	8	Q8WCC1	Q8wcc1 bachia flav	758	7	0.5	217	8	Q9B0I5	Q9B0i5 limnodynast
686	7	0.5	207	8	Q8WCC0	Q8wcc0 procellosau	759	7	0.5	217	16	Q9RIZ6	Q9riz6 streptomyce
687	7	0.5	207	8	Q8WC89	Q8wc89 procellosau	760	7	0.5	218	8	Q9T7Z7	Q9T7z7 lampropelti
688	7	0.5	207	8	Q8WC88	Q8wc88 psilophthal	761	7	0.5	218	8	Q9T7Z6	Q9T7z6 lampropelti
689	7	0.5	207	8	Q8WC87	Q8wc87 calyptommat	762	7	0.5	218	8	Q9T7Z5	Q9T7z5 lampropelti
690	7	0.5	207	8	Q8WC86	Q8wc86 calyptommat	763	7	0.5	218	8	Q9T7Z4	Q9T7z4 lampropelti
691	7	0.5	207	8	Q8WC85	Q8wc85 micrablepha	764	7	0.5	218	8	Q9T7Z3	Q9T7z3 lampropelti
692	7	0.5	207	8	Q8WC84	Q8wc84 bachia bres	765	7	0.5	218	8	Q9T7Z2	Q9T7z2 lampropelti
693	7	0.5	207	8	Q8WC83	Q8wc83 anotosaura	766	7	0.5	218	8	Q9T7Z1	Q9T7z1 lampropelti
694	7	0.5	207	8	Q8WC82	Q8wc82 neusticurus	767	7	0.5	218	8	Q9T7Z0	Q9T7z0 lampropelti
695	7	0.5	207	8	Q8WC81	Q8wc81 placosoma c	768	7	0.5	218	8	Q9T7Y9	Q9T7y9 lampropelti
696	7	0.5	207	8	Q8WC80	Q8wc80 pantodactyl	769	7	0.5	218	8	Q9T7Y8	Q9T7y8 lampropelti
697	7	0.5	207	8	Q8WC89	Q8wc89 colobodactyl	770	7	0.5	218	8	Q9T7Y7	Q9T7y7 lampropelti
698	7	0.5	207	8	Q8WC88	Q8wc88 pantodactyl	771	7	0.5	218	8	Q9T7Y6	Q9T7y6 lampropelti
699	7	0.5	207	8	Q8WC87	Q8wc87 cercosaura	772	7	0.5	218	8	Q9T7Y5	Q9T7y5 lampropelti
700	7	0.5	207	8	Q8WC86	Q8wc86 pholidobolu	773	7	0.5	218	8	Q9T7Y4	Q9T7y4 lampropelti
701	7	0.5	207	8	Q8WC85	Q8wc85 heterodactyl	774	7	0.5	218	8	Q9T7Y3	Q9T7y3 lampropelti
702	7	0.5	207	8	Q8WC84	Q8wc84 colobosau	775	7	0.5	218	8	Q9T7Y2	Q9T7y2 lampropelti
703	7	0.5	207	8	Q8WC83	Q8wc83 colobosau	776	7	0.5	218	8	Q9T7Y1	Q9T7y1 lampropelti
704	7	0.5	207	8	Q8WC82	Q8wc82 tretioscinc	777	7	0.5	218	8	Q9T7Y0	Q9T7y0 lampropelti
705	7	0.5	207	8	Q8WC81	Q8wc81 iphisa eleg	778	7	0.5	218	8	Q9T7X9	Q9T7x9 lampropelti
706	7	0.5	207	8	Q8WC80	Q8wc80 neusticurus	779	7	0.5	218	8	Q9T7X8	Q9T7x8 lampropelti
707	7	0.5	207	8	Q8WC99	Q8wc99 tretioscinc	780	7	0.5	218	8	Q9T7X7	Q9T7x7 lampropelti
708	7	0.5	207	8	Q8WC98	Q8wc98 bachia dorb	781	7	0.5	218	8	Q9T7X6	Q9T7x6 lampropelti
709	7	0.5	207	8	Q8WC97	Q8wc97 prionodactyl	782	7	0.5	218	8	Q9T7X5	Q9T7x5 lampropelti
710	7	0.5	207	8	Q8WC96	Q8wc96 arthrosaura	783	7	0.5	218	8	Q9B3U2	Q9B3u2 plethodon w
711	7	0.5	207	8	Q8WC95	Q8wc95 prionodactyl	784	7	0.5	218	8	Q9B3T5	Q9B3t5 plethodon s
712	7	0.5	207	8	Q8WC94	Q8wc94 prionodactyl	785	7	0.5	218	8	Q9B0J5	Q9B0j5 limnodynast
713	7	0.5	207	8	Q8WC93	Q8wc93 leposoma os	786	7	0.5	218	8	Q9B3U6	Q9B3u6 plethodon s
714	7	0.5	207	8	Q8WC91	Q8wc91 colobosaura	787	7	0.5	218	8	Q9B3V5	Q9B3v5 aneides lug
715	7	0.5	207	8	Q8WC90	Q8wc90 notobachia	788	7	0.5	218	8	Q9T3U7	Q9T3u7 lampropelti
716	7	0.5	207	8	Q8WC78	Q8wc78 tupinambis	789	7	0.5	218	8	Q9T3U6	Q9T3u6 lampropelti
717	7	0.5	207	8	Q8WC77	Q8wc77 kentropyx c	790	7	0.5	218	8	Q9T350	Q9T350 lampropelti
718	7	0.5	207	8	Q8WC76	Q8wc76 cnemidophor	791	7	0.5	218	10	Q93WG8	Q93wg8 oryza sativ
c 719	7	0.5	207	15	Q8UTE4	Q8ute4 human immun	c 792	7	0.5	218	13	Q9FU72	Q9fu72 cynops pyrr
c 720	7	0.5	207	17	Q8TWP9	Q8twp9 methanopyru	c 793	7	0.5	218	16	P96227	P96227 mycobacteri
c 721	7	0.5	208	5	Q8TWD3	Q8twd3 glossina mo	794	7	0.5	219	3	Q12579	Q12579 chaetomium
c 722	7	0.5	208	10	Q9M5W7	Q9m5w7 vaccinium c	795	7	0.5	219	8	Q9MI67	Q9mi67 agkistrodon
723	7	0.5	208	10	Q04033	Q04033 arabidopsis	796	7	0.5	219	8	Q9B3V4	Q9B3v4 plethodon c
c 724	7	0.5	208	16	Q8XDX2	Q8xdx2 escherichia	797	7	0.5	219	8	Q9B0L1	Q9B0l1 limnodynast
c 725	7	0.5	209	5	Q8SQN4	Q8sqn4 encephalito	798	7	0.5	219	8	Q9B0K9	Q9B0k9 limnodynast
726	7	0.5	209	8	Q9B0J4	Q9b0j4 limnodynast	799	7	0.5	219	8	Q9B0K8	Q9B0k8 limnodynast
727	7	0.5	209	8	Q952E1	Q952e1 lineatriton	800	7	0.5	219	8	Q9B0I6	Q9B0i6 limnodynast
728	7	0.5	209	8	Q8WC79	Q8wc79 pantodactyl	801	7	0.5	219	8	Q9B0I4	Q9B0i4 limnodynast
c 729	7	0.5	209	17	Q27345	Q27345 methanobact	802	7	0.5	219	8	Q9B0I2	Q9B0i2 limnodynast
730	7	0.5	210	5	Q9WZ27	Q9wz27 drosophila	803	7	0.5	219	8	Q9B3W0	Q9B3w0 aneides fer
731	7	0.5	210	17	Q9HQC4	Q9hqc4 halobacteri	804	7	0.5	219	8	Q9B3W1	Q9B3w1 aneides fer

c 805	7	0.5	220	1	Q9UXQ2	Q9uxq2 methanobact	878	7	0.5	225	5	Q9NG62	Q9ng62 drosophila
c 806	7	0.5	220	5	Q9N7P5	Q9n7p5 leishmania	879	7	0.5	225	8	Q9B0M5	Q9b0m5 limnodynast
c 807	7	0.5	220	8	Q9B4G9	Q9b4g9 eryx miliar	880	7	0.5	225	8	Q9B0M4	Q9b0m4 limnodynast
c 808	7	0.5	220	8	Q9B3V1	Q9b3v1 plethodon n	881	7	0.5	225	8	Q9B0M3	Q9b0m3 limnodynast
c 809	7	0.5	220	8	Q9B309	Q9b3u9 plethodon o	882	7	0.5	225	8	Q9B0M2	Q9b0m2 limnodynast
c 810	7	0.5	220	8	Q9B3U7	Q9b3u7 plethodon s	883	7	0.5	225	8	Q9B0M1	Q9b0m1 limnodynast
c 811	7	0.5	220	16	Q9CWS8	Q9cms8 pasteurella	884	7	0.5	225	8	Q9B0M0	Q9b0m0 limnodynast
c 812	7	0.5	220	16	Q9K3I0	Q9k3i0 streptomyce	885	7	0.5	225	8	Q9B0L9	Q9b0l9 limnodynast
c 813	7	0.5	220	17	Q9HLA8	Q9hla8 thermoplasm	886	7	0.5	225	8	Q9B0L8	Q9b0l8 limnodynast
c 814	7	0.5	220	17	Q8U207	Q8u207 pyrococcus	887	7	0.5	225	8	Q9B0L7	Q9b0l7 limnodynast
c 815	7	0.5	221	8	Q9MI68	Q9mi68 agkistrodon	888	7	0.5	225	8	Q9B0L6	Q9b0l6 limnodynast
c 816	7	0.5	221	8	Q9B3T0	Q9b3t0 plethodon v	889	7	0.5	225	8	Q9B0L5	Q9b0l5 limnodynast
c 817	7	0.5	221	8	Q9B3V9	Q9b3v9 aneides fla	890	7	0.5	225	8	Q9B3T8	Q9b3t8 plethodon e
c 818	7	0.5	221	12	Q8S299	Q8s299 orf virus	891	7	0.5	225	8	Q9B3T7	Q9b3t7 plethodon e
c 819	7	0.5	221	16	Q985A6	Q985a6 rhizobium l	892	7	0.5	225	8	Q9B3T6	Q9b3t6 plethodon e
c 820	7	0.5	221	16	Q8YKM9	Q8ykm9 anabaena sp	893	7	0.5	225	8	Q9B3T2	Q9b3t2 plethodon n
c 821	7	0.5	221	16	Q9S285	Q9s285 streptomyce	894	7	0.5	225	8	Q9B3T1	Q9b3t1 plethodon n
c 822	7	0.5	222	2	Q9L7R0	Q9l7r0 streptococc	895	7	0.5	225	8	Q9B0L4	Q9b0l4 limnodynast
c 823	7	0.5	222	2	Q9X622	Q9x622 bacillus sp	896	7	0.5	225	8	Q9B0K7	Q9b0k7 limnodynast
c 824	7	0.5	222	5	Q9NLE8	Q9nle8 leishmania	897	7	0.5	225	8	Q9B0K6	Q9b0k6 limnodynast
c 825	7	0.5	222	8	Q9B3U5	Q9b3u5 plethodon i	898	7	0.5	225	8	Q9B0K5	Q9b0k5 limnodynast
c 826	7	0.5	222	8	Q955R6	Q955r6 mictrurus co	899	7	0.5	225	8	Q9B0K4	Q9b0k4 limnodynast
c 827	7	0.5	222	8	Q955R5	Q955r5 mictrurus fr	900	7	0.5	225	8	Q9B0K2	Q9b0k2 limnodynast
c 828	7	0.5	222	8	Q955R4	Q955r4 mictrurus fr	901	7	0.5	225	8	Q9B0K1	Q9b0k1 limnodynast
c 829	7	0.5	222	8	Q955R3	Q955r3 mictrurus al	902	7	0.5	225	8	Q9B0J9	Q9b0j9 limnodynast
c 830	7	0.5	222	8	Q955R2	Q955r2 mictrurus al	903	7	0.5	225	8	Q9B0J8	Q9b0j8 limnodynast
c 831	7	0.5	222	8	Q955R1	Q955r1 mictrurus al	904	7	0.5	225	8	Q9B0J6	Q9b0j6 limnodynast
c 832	7	0.5	222	8	Q955R0	Q955r0 mictrurus al	905	7	0.5	225	8	Q9B0J3	Q9b0j3 limnodynast
c 833	7	0.5	222	8	Q955Q9	Q955q9 mictrurus ba	906	7	0.5	225	8	Q9B0J2	Q9b0j2 limnodynast
c 834	7	0.5	222	8	Q955Q8	Q955q8 mictrurus py	907	7	0.5	225	8	Q9B0J1	Q9b0j1 limnodynast
c 835	7	0.5	222	8	Q955Q7	Q955q7 mictrurus le	908	7	0.5	225	8	Q9B0J0	Q9b0j0 limnodynast
c 836	7	0.5	222	8	Q955Q6	Q955q6 mictrurus le	909	7	0.5	225	8	Q9B0I8	Q9b0i8 limnodynast
c 837	7	0.5	222	8	Q955Q5	Q955q5 mictrurus le	910	7	0.5	225	8	Q9B0I7	Q9b0i7 limnodynast
c 838	7	0.5	222	8	Q955Q4	Q955q4 mictrurus le	911	7	0.5	225	8	Q9B0I3	Q9b0i3 limnodynast
c 839	7	0.5	222	8	Q955Q3	Q955q3 mictrurus le	912	7	0.5	225	8	Q9B0I1	Q9b0i1 limnodynast
c 840	7	0.5	222	8	Q955Q2	Q955q2 mictrurus ib	913	7	0.5	225	8	Q9B3V6	Q9b3v6 aneides lug
c 841	7	0.5	222	8	Q955Q1	Q955q1 mictrurus de	914	7	0.5	225	8	Q9B3V8	Q9b3v8 aneides fla
c 842	7	0.5	222	8	Q955Q0	Q955q0 mictrurus he	915	7	0.5	225	12	Q9BWZ5	Q9bwz5 shope fibro
c 843	7	0.5	222	8	Q955P9	Q955p9 mictrurus sp	c 916	7	0.5	226	8	Q953I5	Q953i5 echinosorex
c 844	7	0.5	222	8	Q955P8	Q955p8 mictrurus su	917	7	0.5	226	8	Q9B3W3	Q9b3w3 ensatina es
c 845	7	0.5	222	8	Q94P67	Q94p67 mictrurus br	918	7	0.5	226	8	Q9B3W4	Q9b3w4 ensatina es
c 846	7	0.5	222	16	Q9HWM7	Q9hwm7 pseudomonas	c 919	7	0.5	226	8	Q8SJ12	Q8sj12 ursus arcto
c 847	7	0.5	223	4	Q9H687	Q9h687 homo sapien	c 920	7	0.5	226	8	Q8SJH0	Q8sjh0 ursus marit
c 848	7	0.5	223	8	Q9B0L3	Q9b0l3 limnodynast	c 921	7	0.5	226	16	Q9I1O3	Q9i1o3 pseudomonas
c 849	7	0.5	223	8	Q94Y31	Q94y31 crotalus vi	922	7	0.5	227	3	Q8X0K5	Q8x0k5 neurospora
c 850	7	0.5	223	8	Q94Y30	Q94y30 crotalus vi	c 923	7	0.5	227	5	Q9BL42	Q9bl42 caenorhabdi
c 851	7	0.5	223	8	Q94Y29	Q94y29 crotalus vi	c 924	7	0.5	227	5	Q9VGF4	Q9vgf4 drosophila
c 852	7	0.5	223	8	Q94Y28	Q94y28 crotalus vi	c 925	7	0.5	227	8	O2I346	O2i346 euhadra her
c 853	7	0.5	223	8	Q94Y27	Q94y27 crotalus vi	926	7	0.5	227	8	Q9B3V7	Q9b3v7 aneides har
c 854	7	0.5	223	8	Q94Y26	Q94y26 crotalus vi	c 927	7	0.5	227	11	Q9CYL2	Q9cyl2 mus musculu
c 855	7	0.5	223	8	Q94Y25	Q94y25 crotalus vi	c 928	7	0.5	227	16	Q9CD41	Q9cd41 mycobacteri
c 856	7	0.5	223	8	Q94Y24	Q94y24 crotalus vi	c 929	7	0.5	227	16	Q8XIA0	Q8xia0 clostridium
c 857	7	0.5	223	8	Q94Y23	Q94y23 crotalus vi	c 930	7	0.5	228	3	O74463	O74463 schizosacch
c 858	7	0.5	223	8	Q94Y22	Q94y22 crotalus vi	931	7	0.5	228	8	Q9B3W5	Q9b3w5 ensatina es
c 859	7	0.5	223	8	Q94Y21	Q94y21 crotalus vi	932	7	0.5	228	8	Q9T7X4	Q9t7x4 coluber con
c 860	7	0.5	223	8	Q94Y20	Q94y20 crotalus vi	933	7	0.5	228	16	Q9RW00	Q9rw00 deinococcus
c 861	7	0.5	223	8	Q94Y19	Q94y19 crotalus vi	c 934	7	0.5	228	10	Q94H60	Q94h60 oryza sativ
c 862	7	0.5	223	8	Q94Y18	Q94y18 crotalus vi	c 935	7	0.5	228	11	Q9CYL8	Q9cyl8 mus musculu
c 863	7	0.5	223	8	Q94Y17	Q94y17 crotalus vi	c 936	7	0.5	228	11	Q9CPQ2	Q9cpq2 mus musculu
c 864	7	0.5	223	8	Q94Y16	Q94y16 crotalus vi	c 937	7	0.5	228	11	Q920H9	Q920h9 rattus norv
c 865	7	0.5	223	8	Q94Y15	Q94y15 crotalus vi	938	7	0.5	228	16	Q9RW00	Q9rw00 deinococcus
c 866	7	0.5	223	8	Q94Y14	Q94y14 crotalus vi	939	7	0.5	228	16	Q9CDY8	Q9cdy8 lactococcus
c 867	7	0.5	223	8	Q94Y13	Q94y13 crotalus vi	940	7	0.5	229	8	Q952E6	Q952e6 lineatriton
c 868	7	0.5	223	8	Q94Y12	Q94y12 crotalus sc	941	7	0.5	229	8	Q8WA06	Q8wa06 dendroaspis
c 869	7	0.5	223	8	Q94Y11	Q94y11 crotalus du	942	7	0.5	229	8	Q8W9Z7	Q8w9z7 ophiophagus
c 870	7	0.5	223	10	Q9W5C3	Q9w5c3 solanum cha	c 943	7	0.5	229	10	Q9LFG5	Q9lpg5 arabiadopsis
c 871	7	0.5	223	11	Q63211	Q63211 rattus norv	c 944	7	0.5	229	16	O8YLP4	O8ylp4 anabaena sp
c 872	7	0.5	224	8	Q9MLQ4	Q9mlq4 limulus pol	c 945	7	0.5	229	17	O8ZW30	O8zw30 pyrobaculum
c 873	7	0.5	224	8	Q9B3T9	Q9b3t9 plethodon e	c 946	7	0.5	230	5	Q9TZ76	Q9tz76 caenorhabdi
c 874	7	0.5	224	8	Q9B0L2	Q9b0l2 limnodynast	947	7	0.5	230	8	Q9MEB5	Q9meb5 agkistrodon
c 875	7	0.5	224	8	Q9B3W6	Q9b3w6 ensatina es	948	7	0.5	230	8	Q9MI71	Q9mi71 agkistrodon
c 876	7	0.5	224	8	Q8W9Z3	Q8w9z3 walterinnes	949	7	0.5	230	8	Q9MI66	Q9mi66 agkistrodon
c 877	7	0.5	224	16	Q92JH1	Q92jh1 rickettsia	950	7	0.5	230	8	Q9MI76	Q9mi76 crotalus ti

951 7 0.5 230 8 Q9M175
952 7 0.5 230 8 Q9M174
953 7 0.5 230 8 Q9M173
954 7 0.5 230 8 Q9M170
955 7 0.5 230 8 Q9M169
956 7 0.5 230 8 Q9M163
957 7 0.5 230 8 Q94UE0
958 7 0.5 230 8 Q8W925
959 7 0.5 230 8 Q8SKS0
960 7 0.5 230 8 Q8SKR7
961 7 0.5 231 8 Q9MDM6
962 7 0.5 231 8 Q79627
963 7 0.5 231 8 Q79628
964 7 0.5 231 8 Q79629
965 7 0.5 231 8 Q79630
966 7 0.5 231 8 Q79631
967 7 0.5 231 8 Q79632
968 7 0.5 231 8 Q79633
969 7 0.5 231 8 Q79472
970 7 0.5 231 8 Q79474
971 7 0.5 231 8 Q79475
972 7 0.5 231 8 Q79476
973 7 0.5 231 8 Q79477
974 7 0.5 231 8 Q79478
975 7 0.5 231 8 Q79479
976 7 0.5 231 8 Q79480
977 7 0.5 231 8 Q79482
978 7 0.5 231 8 Q79483
979 7 0.5 231 8 Q79484
980 7 0.5 231 8 Q79485
981 7 0.5 231 8 Q79487
982 7 0.5 231 8 Q79489
983 7 0.5 231 8 Q79491
984 7 0.5 231 8 Q79492
985 7 0.5 231 8 Q79493
986 7 0.5 231 8 Q79494
987 7 0.5 231 8 Q79496
988 7 0.5 231 8 Q79497
989 7 0.5 231 8 Q79498
990 7 0.5 231 8 Q79500
991 7 0.5 231 8 Q79501
992 7 0.5 231 8 Q79502
993 7 0.5 231 8 Q79503
994 7 0.5 231 8 Q9T621
995 7 0.5 231 8 Q9T620
996 7 0.5 231 8 Q9T619
997 7 0.5 231 8 Q9T618
998 7 0.5 231 8 Q9T617
999 7 0.5 231 8 Q9T616
1000 7 0.5 231 8 Q9T7X3

ALIGNMENTS

RESULT 1
Q96KF5 PRELIMINARY; PRT; 1320 AA.
AC Q96KF5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Myopalladin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]

RN SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE:
RX MEDLINE=21206024; PubMed=11309420;
RA Bang M.L., Mudry R.E., McElhinny A.S., Trombitas K., Geach A.J.,
RA Yamasaki R., Sorimachi H., Granzier H., Gregorio C.C., Labeit S.;
RT "Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple

RT roles in 2-disc and I-band protein assemblies.";
RL J. Cell Biol. 153:413-428(2001).
DR EMBL; AF328296; AAK50625.1; -.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 5.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
SQ SEQUENCE 1320 AA; 145379 MW; A6579FB164D33B6E CRC64;

Alignment Scores:

Pred. No.: 0 Length: 1320
Score: 770.00 Matches: 1314
Percent Similarity: 99.24% Conservative: 0
Best Local Similarity: 99.24% Mismatches: 6
Query Match: 58.51% Indels: 8
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q96KF5 (1-1320)

Qy 1 ATGCAAGACGACGACATAGAAAGCTTCTACTTCCATATCTCAGCTTCTAAGAGAGAGCTAT 60
Db 1 MetGlnAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
Qy 61 TTAGCTGAACACGACATCGGGAAACAATGAGAGAGTCGAGCGGAGCCCTCTCTCAAC 120
Db 21 LeuAlaGluThrArgHisArgGlyAsnAsnGluArgSerArgAlaGluProSerSerAsn 40
Qy 121 CCTTGCCATTTCGGCAGCTCTCTGGGGCGCTCAAGGAGCGGAGCCCAAGATGACCTT 180
Db 41 ProcysHisPheGlySerProSerGlyAlaGluGlyGlyGlyGlnAspAspLeu 60
Qy 181 CCAGATCTTTTCAGCTTTCGAGCCAAAGAAATAGACGAAAGTGTCAATTTGGCAAGA 240
Db 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
Qy 241 CTGGCCATCAATTACGACCTTTGGAGAGCGAGATGAAACTCAAGCTAGAAAACGACTT 300
Db 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
Qy 301 TCTCTCATCAGATGAACACTCACCTAATTTAGTTTTCAGCTTAACCTTCGCCAGAT 360
Db 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
Qy 361 AACCTCGAAGTCCACAGCTCTAAAGAAAGCCCCAGGAGGCAAAAGCCACAGTAT 420
Db 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
Qy 421 TGTCTGAAACCCAGTCCAAAAGATATTTTAAATAGGCTGCCGACTTCATTGAAGAG 480
Db 141 CysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
Qy 481 CTATCTCTCCCTTTCAAAATCCACAGCTCCAAAAGGATTAGACCTCGTGCCTCAAAAC 540
Db 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
Qy 541 CACAAGAGTAACTGGAATCTCAAAACAAAGTTATGTCAGGAAACAGCTCCAGCTTTCTCA 600
Db 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
Qy 601 GATCTGTCAAGACGAGAAAGATCTTCTGTTCCCATCCCTATCCCTGCGGATACACAG 660
Db 201 AspLeuSerGluArgArgGluArgSerSerValProIleProIleProAlaAspThrArg 220
Qy 661 GATAATCAAGTGAATCAGCCCTCGGAACACGACGAGCCCAAGAGCGCTGAAGCGGACAG 720
Db 221 AspAsnGluValAsnHisAlaLeuGluGlnGlnGluAlaLysArgArgGluAlaGluGln 240
Qy 721 GCTGCCAGTGGCGGCTGGTGGAGACACTACACAGGGCTTTCCCTTCATCTCTGTAC 780
Db 241 AlaAlaSerGluAlaAlaGlyGlyAspThrThrProGlySerSerProSerSerLeuTyr 260
Qy 781 TATGAAGAACTCTGGGGCAACCTCCCGGTTCTACTCAAAAGTTACGGACGAGAGATT 840
Db 141

Db 261 TyrGluGluProLeuGluGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
QY 841 CCAGAAGGAACGAGTACAGTTGGATTGCATAGTGGTAGGAATCCACACCTCAAGTA 900
Db 281 ProGluGlyThrArgValGlnLeuAspCysIleValValGlyIleProProGlnVal 300
QY 901 AGGTGGTACTGTGAAGCGAAGGAGCTTGAAATTCGCCAGATATTCACATCGTCCAGGCA 960
Db 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
QY 961 GGAATCTGCACCTACCTGACCATTGGGAAGCCTTTGAAGAGGACACAGACGCTATTCC 1020
Db 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
QY 1021 TGTCTTTGCTTCTAACATCATGGCAGAGATTCGACTTCTGCTCAGATTTATATAGAAGG 1080
Db 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
QY 1081 GTTCTTCTTCTGACTCAGAAGCGACCCCTTAACAAGGAAGAGATGAATCGAATCCAGAAG 1140
Db 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
QY 1141 CCAATAGAGTGCATCTCTCCCATCTCTCCCATCTCTGCGAGTCAATTCCTCAGCAGTACCCCAA 1200
Db 381 ProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValProGln 400
QY 1201 GCCCAGCATTTGGTGCCCAACCTCGTGTGGCAACCATCCAGCAGTGTACAGACCCACC 1260
Db 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
QY 1261 AATTACTTGCAGGGATTGGATGGAACCTATCATTTGCAGCTCCTGTGTTTACAAAGATG 1320
Db 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleAlaAlaProValPheThrLysMet 440
QY 1321 CTCAAAATTTGTCAGTCTTGAGGTCAGCTGTTGTCTTTGAATGCAGAGTAAAGGA 1380
Db 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
QY 1381 GCTCCATCTCTTAAGTTGAGTGGTATAGAGAAGGACTTTAATAGAAGATTCTCCAGAT 1440
Db 461 AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
QY 1441 TTTAGGATTTACAGAAAAACCTCGATCCATCGCAGAGCAGAGGAGATTTGCACCTTG 1500
Db 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluProGluGluIleCysThrLeu 500
QY 1501 GTCATTGCTGAGTGTTCGAGAGATTCTGGGTGCTTCATCTGACTGCTCAACACACAAA 1560
Db 501 ValIleAlaGluValPheAlaGluAspSerGlyCysPheThrCysThrAlaSerAsnLys 520
QY 1561 TACGGCAGCTGTCAAGCATTCACAGCTGCAGTGCAGAGGAAATGAGGACCTCAGCAAC 1620
Db 521 TyrGlyThrValSerSerIleAlaGlnLeuHisValArgGlyAsnGluAspLeuSerAsn 540
QY 1621 AACGGGTCTCTTCACCTCAGCAACTCYACACCACCACTGGCAGCTATTGAGCCACAGCCC 1680
Db 541 AsnGlySerLeuHisSerAlaAsnSerThrThrAsnLeuAlaAlaIleGluProGlnPro 560
QY 1681 TCCCCACCCACACTCAGAGCCTCATCTGTGGAAACACACCCCAACCCCAACCTCAGGGG 1740
Db 561 SerProHisSerGluProProSerValGluGlnProProLysProLysLeuGluGly 580
QY 1741 GTTCTGGTGAACACATGAGCCCGCTCCAGCTCCAGGATTGGCTTCGTGTGCACCTTC 1800
Db 581 ValLeuValAsnHisAsnGluProArgSerSerSerArgIleGlyLeuArgValHisPhe 600
QY 1801 AACCTGCTGAAGATGACAAAGGAAGTGAAGCATCTCCAGGCTGGTGGTGAACACC 1860
Db 601 AsnLeuProGluAspAspLysGlySerGluAlaSerSerGluAlaGlyValValThrThr 620
QY 1861 AGACAGACAGGCCCGATTCTTTSCAGGAGAGGTTCAACGGACAGACACAAAAACCCC 1919
Db 621 ArgGlnThrArgProAspSerPhe-GlnGluArgPheAsnGlyGlnAlaThrLysThrPr 640

QY 1920 AGAGCCTTCTTTCCCC - GTGAAAGAGCCCTCCAGTCTTGCCAAACCCAAACTTGATT 1978
Db 640 OGLuProSer - SerProValLysGluProProValLeuAlaLysProLysLeuAspS 660
QY 1979 CCACTCAGTTACACAGCTTCATAACCAAGTCTTACTGGAACCAACCAATTCAAAACC 2038
Db 660 erThrGlnLeuGlnLeuHisAsnGlnValLeuLeuGluGlnHisGlnLeuGlnAsnPr 680
QY 2039 CACCTCCTTCATCTCCTAAGGAGTTTCCCTTTCARCATGACTGTTTGAACCTCAATGCTC 2098
Db 680 roProProSerSerProLysGluPheProPheAsnMetThrValLeuAsnSerAsnAlaP 700
QY 2099 CCCACGCGGTGACAACTCCARTAAGCAGGTGAAGGCTCCTTCATCAGACGCTTCAGCT 2158
Db 700 roProAlaValThrThrSerAsnLysGlnValLysAlaProSerSerGlnThrPheSerL 720
QY 2159 TGCCCGCGCCGGAAGTATTTCCTCCCTCCACGAACACCACCGCAGCAACTGTGGCCCTT 2218
Db 720 euAlaArgProLysTyrPhePheProSerThrAsnThrThrAlaAlaThrValAlaProS 740
QY 2219 CCAGCTCTCCGGTGTTCACCTTTGAGCAGCAGCTCCTCAAACATATTCAGAGACAGTGAC 2278
Db 740 erSerSerProValPheThrLeuSerSerThrProGlnThrIleGlnArgThrValSerL 760
QY 2279 AAGAAAGCCTCTTAGTGTCTCACCCCTCTGTGCAAAACCAATCTCCAGGAGGCTTTCCA 2338
Db 760 ysluSerLeuLeuValSerHisProSerValGlnThrLysSerProGlyLysLeuSerI 780
QY 2339 TCCAAATGAGCCACTCCACACGAGGCCAACAGAACACACACACCACTTCACATTTT 2398
Db 780 leGlnAsnGluProLeuProProGlyProThrGluProThrProProPheThrPheS 800
QY 2399 CCATCCCCAGC - GGAACACAGTTTCAGCCCGCTGTGTGTCCTCCCAATTCCTGTCTCCT 2457
Db 800 erIle - ProArgGlyAsnGlnPheGlnProArgCysValSerProIleProValSerPro 819
QY 2458 ACCAGCGGATTTCAGAACCCAGTGGCTTCTCAGCTCTGCTGCTTCTCCTCCTCCCTGCC 2517
Db 820 ThrSerArgIleGlnAsnProValAlaPheLeuSerSerValLeuProSerLeuProAla 839
QY 2518 ATCCCCACCAAAATGCCAT - GGGCTGCTAGAGTGCACCATCCATCCATCCACAGG 2576
Db 840 ileProProThrAsnAlaMetGly - LeuProArgSerAlaProSerMetProSerGlnGl 859
QY 2577 ATTAGCGAAGAAAAATACAAAGTCTCCTCAACAGTGAATGATGATACATTCGTGNAAC 2636
Db 859 yLeuAlaLysLysAsnThrLysSerProGlnProValAsnAspAsnIleArgGluTh 879
QY 2637 TAAGAACGCACTGATTCGAGACTTGGGAAAAATAAATTCAGTGATGTCAGACCAAA 2696
Db 879 rLysAsnAlaValIleArgAspLeuGlyLysLysIleThrPheSerAspValArgProAs 899
QY 2697 CCACGAGGAGTACAAAATTTCAAGCTTTGACGAGAGGCTGATGAATGAATAGAGTTTCG 2756
Db 899 nGlnGlnLysLysIleSerSerPheGluGlnArgLeuMetAsnGluIleGluPheAr 919
QY 2757 CTTGGACGCTACTCTCTGTTGATGAATCAGATGATGAATTCACATGATGAGATCCCCAC 2816
Db 919 gLeuGluArgThrProValAspGluSerAspAspGluIleGlnHisAspGluIleProTh 939
QY 2817 GGCACAGTGTATTCCTCCCATCTTGCACAGAGACTCAACACTTCCTCGGCTCACAGAAG 2876
Db 939 rGlyLysCysIleAlaProIlePheAspLysArgLeuLysHisPheArgValThrGluGl 959
QY 2877 CTCTCAGTCACATTCACCTGCAAAATTTGGGTACCTCTGTTCCAAAGGTTTACTGGT 2936
Db 959 ySerProValThrPheThrCysLysIleValGlyIleProValProLysValTyrTrpPh 979
QY 2937 CAAAGATGGGAAGCAGATTTCTAAGAAATAGCACTGCAAAATAGAGCGAGAGAGAGA 2996
Db 979 eLysAspGlyLysGlnIleSerLysArgAsnGluHisCysLysMetArgArgGluGlyAs 999

```
QY 2997 TGGGACATGCTCTCGACATTTGAATCCATACACATGATGACGATGGCAGCACTACACCAT 3056
DB 999 pGlyThrCysSerLeuHisIleGluSerThrThrSerAspAspGlyAsnTyrThrIle 1019
QY 3057 CATGGCAGCCAAACCCAGGGGAGAATCAGCTGTTCTGGCCACTTGATGGTACAAAGTTT 3116
DB 1019 eMetAlaAlaAsnProGlnGlyArgIleSerCysSerGlyHisLeuMetValGlnSerLe 1039
QY 3117 GCCCATTCGCGAGTCGGGTCAACTCTGCTGGTTCAGTCTCAGAGGGGAAGATCCCGAGTGCA 3176
DB 1039 uProIleArgSerArgLeuThrSerAlaGlyGlnSerHisArgGlyArgSerArgValGI 1059
QY 3177 AGAAGAGACAAAGAGCCCTACAGAACGCTTTTCCGACCAATTCCTGCGAGGCTCC 3236
DB 1059 nGluArgAspLysGluProLeuGlnGluArgPhePheArgProHisPheLeuGlnAlaPr 1079
QY 3237 TGGGGATATGGTACTCATGAGGGGCCCTCTGCGCTGGACTGTAAAGTGAGTGTTT 3296
DB 1079 oGlyAspMetValAlaHisGluGlyArgLeuCysArgLeuAspCysLysValSerGlyLe 1099
QY 3297 ACCGCCCCCGGAGCTGACATGGCTTACTCAATGGCCAAACCTGTGTACACGATGCTCCCA 3356
DB 1099 uProProGluLeuThrTrpLeuLeuAsnGlyGlnProValLeuProAspAlaSerHi 1119
QY 3357 CAAGATGCTGTGAGGAGACCGGAGTCCACTCTCTGCTCATTTGACCACTCACTCAGCG 3416
DB 1119 sLysMetLeuValArgGluThrGlyValHisSerLeuLeuIleAspProLeuThrGlnAr 1139
QY 3417 CGACGCGAGGACCTTAAATGTCATCGCTACCAACAAACCGGCGAGATTCCTTTAGTCT 3476
DB 1139 gAspAlaGlyThrTrpLysCysIleAlaThrAsnLysThrGlyGlnAsnSerPheSerLe 1159
QY 3477 GGAGCTCTCTGTAGTAGCAAAAGAGGTGAAGAAAGCACTGTGATCTGTGAGAAACTACA 3536
DB 1159 uGluLeuSerValValAlaLysGluValLysLysAlaProValIleLeuGluLysLeuGI 1179
QY 3537 GAATCGCGGTGTTCCCGAAGGCCACCCCGTGAGACTCGAGTGGCGCGGTAGGCAATGCC 3596
DB 1179 nAsnCysGlyValProGluGlyHisProValArgLeuGluCysArgValIleGlyMetPr 1199
QY 3597 CCCACCTGTGTTCTTACTGGAAGAAGACAATGAGACCATCCCTGCGACGAGAGAGAT 3656
DB 1199 oProProValPheTrpLysLysAspAsnGluThrIleProCysThrArgGluArgII 1219
QY 3657 CAGTATGACGACGACACACAGGATGCTGCTGCTCTCTCATTCAGCCAGCAAGAAATC 3716
DB 1219 eSerMetHisGlnAspThrThrGlyTyrAlaCysLeuLeuIleGlnProAlaLysLysSe 1239
QY 3717 AGACGCTGGATGGTACAGTTGTGACGCCAAGATGAAGCGCGCATCGTGTGCGCACTGC 3776
DB 1239 rAspAlaGlyTrpTrpThrLeuSerAlaLysAsnGluAlaGlyIleValSerCysThrAl 1259
QY 3777 CAGGCTGATATATACGCTCAGTGGCAACATGAGACATCCACCGCCCATGCTGTCGCGCC 3836
DB 1259 aArgLeuAspIleTrpAlaGlnTrpHisGlnIleProProMetSerValArgPr 1279
QY 3837 CAGTGGCAGCTGCTACGAGTCTCTCACCAGTAAAGGACTTGACATATTTCTGCTTTTC 3896
DB 1279 oSerGlySerArgTyrGlySerLeuThrSerLysGlyLeuAspIlePheSerAlaPheSe 1299
QY 3897 CTCCATGGAAGACGATGGTGATTTTCATGCTTCTCGGAGTGATGGAGAGTGATGA 3956
DB 1299 rSerMetGluSerThrMetValTyrSerCysSerArgSerValValGluSerAspGI 1319
QY 3957 ACTT 3960
DB 1319 uLeu 1320

RESULT 2
Q96K90 PRELIMINARY; PRT; 507 AA.
AC Q96K90;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
```

```
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE CDNA FLJ14437 fis, clone HEMBB1000083, weakly similar to myosin light
DE chain kinase, smooth muscle and NON-muscle isozymes
DE (EC 2.7.1.117).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DE EMBL: AK027343; BAB55048.1;
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF000047; Ig; 2.
SQ SEQUENCE 507 AA; 56185 MW; BD484878845DF39 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 507
Score: 494.00 Matches: 494
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.54% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q96K90 (1-507)

QY 1 ATCAAGACGACAGCATAGAGCTTCTACTTCCATATCTCAGCTTCTTAAGAGAGAGCTAT 60
DB 1 MetGlnAspAspSerIleGluAlaSerThrSerIleSerGlnLeuLeuArgGluSerTyr 20
QY 61 TTAGCTAAACCCAGACATCGGGGAAACAATGAGAGGAGTCGAGGGAGCCCTCTCTCAAC 120
DB 21 LeuAlaGluThrArgHisArgGlyAsnGluArgSerArgAlaGluProSerSerAsn 40
QY 121 CCTTGCCATTTCGGCAGCTCCTCTCTGGGCGCGCTGAAGAGCGGAGGCAAGATGACCTT 180
DB 41 ProCysHisPheGlySerProSerGlyAlaAlaGluGlyGlyGlyGlnAspAspLeu 60
QY 181 CCAGATCTTTTCAGCCTTTCTGAGCCAAAGAAGATTAGACGAAAGTGTCAATTTGGCAAG 240
DB 61 ProAspLeuSerAlaPheLeuSerGlnGluLeuAspGluSerValAsnLeuAlaArg 80
QY 241 CTGGCCATCAATTACGACCCCTTTGGAGAAGCGAGATCAAACTCAAGCTAGAAAACGACTT 300
DB 81 LeuAlaIleAsnTyrAspProLeuGluLysAlaAspGluThrGlnAlaArgLysArgLeu 100
QY 301 TCTCCTCATCAGATGAACACACTCACCTTAATTTAAGTTTTCAGCCTTAATCTCTCCAGGAT 360
DB 101 SerProAspGlnMetLysHisSerProAsnLeuSerPheGluProAsnPheCysGlnAsp 120
QY 361 AACCTCGAAGTCCCACTCCAGCTCTTAAAGAAAGCCCGAGAGGCAAAAGCCACAGTAT 420
DB 121 AsnProArgSerProThrSerSerLysGluSerProGlnGluAlaLysArgProGlnTyr 140
QY 421 TGTCTCTGAAACCCAGTCCAAAAAGATATTTTAAATAAGGCTGCCGACTTCATTGAGAG 480
DB 141 CysSerGluThrGlnSerLysValPheLeuAsnLysAlaAlaAspPheIleGluGlu 160
QY 481 CTATCTCTCCCTTTTCAATCCACAGCTCCAAAAGGATTAGACTCTGCTGCTGCTCAAAAC 540
DB 161 LeuSerSerLeuPheLysSerHisSerLysArgIleArgProArgAlaCysLysAsn 180
QY 541 CACAAGAGTAACTGGGAATCTCAAAACAAAGATTATGACGAAACACAGCTCCAGCTTCTCA 600
```

```
Db 181 HisLysSerLysLeuGluSerGlnAsnLysValMetGlnGluAsnSerSerPheSer 200
|||
QY 601 GATCTGTCACAAAGACGAGAAATCTTCGTTCCCATCCTATCCCTCGGGATACAGG 660
|||
Db 201 AspLeuSerGluArgGluArgSerValProIleProIleProAlaAspThrArg 220
|||
QY 661 GATAATGAAGTGAATCACGCCCTCGAACACAGCAGAGCCAGAGGGCTGAAGGGAGCAG 720
|||
Db 221 AspAsnGluValAsnHisAlaLeuGluGlnGluAlaLysArgArgGluAlaGluGln 240
|||
QY 721 GCTGCCAGTCAGGGCTGTGTGAGACACATACACAGGGTCTCCCTTCATCTCTGTAC 780
|||
Db 241 AlaAlaSerGluAlaAlaGlyGlyAspThrProGlySerSerProSerSerLeuTyr 260
|||
QY 781 TATGAAGAACCTCTGGGGCAACCTCCCGGTTTCACATCAAAAGTTACGGACGAGAGTT 840
|||
Db 261 TyrGluGluProLeuGluGlnProProArgPheThrGlnLysLeuArgSerArgGluVal 280
|||
QY 841 CCAGAAGAACTCGAGTACAGTGTGGATTGCATAGTGTAGGAATTCACACACCTCAAGTA 900
|||
Db 281 ProGluGlyThrArgValGlnLeuAspCysIleValGlyIleProProGlnVal 300
|||
QY 901 AGTGTGTACTGTGAAGCAAGAGAGCTTGAAATTTCCACAGATATTCATCGTCCAGGCA 960
|||
Db 301 ArgTrpTyrCysGluGlyLysGluLeuGluAsnSerProAspIleHisIleValGlnAla 320
|||
QY 961 GGAATCTGCACCTGACCATTCGGAGAGCTTTCAGAGACACAGACGCTATTC 1020
|||
Db 321 GlyAsnLeuHisSerLeuThrIleAlaGluAlaPheGluGluAspThrGlyArgTyrSer 340
|||
QY 1021 TGCTTTCTCTTAACATCTATGGACAGATTCGACTTCTCTGAGATTTATATAGAGGG 1080
|||
Db 341 CysPheAlaSerAsnIleTyrGlyThrAspSerThrSerAlaGluIleTyrIleGluGly 360
|||
QY 1081 GTTCTTCTTCTGACTCAGAAGCGCCCTTAACAAGGAAGATGAATCAATCCAGAAG 1140
|||
Db 361 ValSerSerSerAspSerGluGlyAspProAsnLysGluGluMetAsnArgIleGlnLys 380
|||
QY 1141 CCAATCAGCTGTCATCTCCCTCCACCTACCTCTGCAGTCACTCTCCAGCAGTACCCAA 1200
|||
Db 381 ProAsnGluValSerSerProProThrThrSerAlaValIleProProAlaValProGln 400
|||
QY 1201 GCCCAGCATTTGTTGGCCCAACCTCTGTGGCAACCATCCAGCAGTGTCCAGAGCCCA 1260
|||
Db 401 AlaGlnHisLeuValAlaGlnProArgValAlaThrIleGlnGlnCysGlnSerProThr 420
|||
QY 1261 AATTACTTCAGGATTTGGATGGAAACCTATCATTCAGCTCTCTGTGTTTCAAAAGATG 1320
|||
Db 421 AsnTyrLeuGlnGlyLeuAspGlyLysProIleIleAlaProValPheThrLysMet 440
|||
QY 1321 CTACAAATTTGTCAGCTCTGAGGTCAGCTGCTTCTTGAATCCAGAGTAAAGGA 1380
|||
Db 441 LeuGlnAsnLeuSerAlaSerGluGlyGlnLeuValValPheGluCysArgValLysGly 460
|||
QY 1381 GCCTCAATCTCTTAAGTTGAGTGTATAGAGAAGGAGCTTTAATAGAGATTTCTCCAGAT 1440
|||
Db 461 AlaProSerProLysValGluTrpTyrArgGluGlyThrLeuIleGluAspSerProAsp 480
|||
QY 1441 TTATAGATTTTACAGAAAAACCTTCGATCCATCCATCCAGAGCCCA 1482
|||
Db 481 PheArgIleLeuGlnLysLysProArgSerMetAlaGluPro 494
|||
RESULT 3
Q9UQF5 PRELIMINARY; PRT; 385 AA.
AC Q9UQF5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE SIH002 (Similar to FALLADIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu T., Zhang J., Ye M., Zhang Q., Fu G., Zhou J., Wu J., Shen Y.,
RA Yu M., Chen S., Mao M., Chen Z.;
RT "Human SIH002 gene.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077041; AAD27774.1; -.
DR EMBL; BC013867; AAH13867.1; -.
DR HSSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00408; IGC2; 3.
KW Immunoglobulin domain.
SQ SEQUENCE 385 AA; 42932 MW; 15853FB78B8E41B9 CRC64;

Alignment Scores:
Pred. No.: 1.76e-10 Length: 385
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.44% Indels: 0
DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9UQF5 (1-385)
QY 3034 GATGACATGGCACTACATCATCGCAGCCAAACCCAGGGAGAAATCAGCTGT 3090
|||||
Db 70 AspAspAspGlyAsnTyrThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 88
|||||
RESULT 4
Q9Y3E9 PRELIMINARY; PRT; 404 AA.
AC Q9Y3E9; Q9UGA0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CGI-151 protein.
GN DXFZP586L0518.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20212150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of Novel Human Genes Evolutionarily Conserved in
Caenorhabditis elegans by Comparative Proteomics.";
RL Genome Res. 10:703-713(2000).
[2]
RN SEQUENCE OF 260-404 FROM N.A.
RP TISSUE=UTERUS;
RC Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF151909; AAD34146.1; -.
DR EMBL; AL050093; CAB43265.1; -.
DR HSSP; P56276; ITLK.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00408; IGC2; 3.
KW Hypothetical protein; Immunoglobulin domain.
SQ SEQUENCE 404 AA; 45005 MW; 729696C78FFC2B5F CRC64;

Alignment Scores:
```

Pred. No.: 1.75e-10 Length: 404
 Score: 19.00 Matches: 19
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.44% Indels: 0
 DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9Y3E9 (1-404)

QY 3034 GATGACGATGCGCACTACACCATCATGCGCAACCCCGAGGGAGAAATCAGCTGT 3090
 Db 89 AspAspAspGlyAsnTyThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 107

RESULT 5

Q9Y2J6 PRELIMINARY; PRT; 772 AA.
 AC Q9Y2J6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE KIAA0992 protein (Fragment).
 GN KIAA0992
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;

Sequence FROM N.A.

TISSUE=Brain;

RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,

RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIII.

RT The complete sequences of 100 new cDNA clones from brain which code

for large proteins in vitro.;"

RL DNA Res. 6:63-70(1999).

DR EMBL; AB023209; BAA76836.1; -

DR HSP; P56276; IYTK.

DR InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR002965; P_rich_extensn.

DR Pfam; PF00047; Ig; 3.

DR PRINTS; PR01217; PRICHEXTENS.

DR SMART; SM00408; IGC2; 3.

KW Immunoglobulin domain.

FT NON_TER 1

SQ SEQUENCE 772 AA; 83643 MW; EE5EB2BC79C7492B CRC64;

Alignment Scores:

Pred. No.: 1.6e-10 Length: 772

Score: 19.00 Matches: 19

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.44% Indels: 0

DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9Y2J6 (1-772)

QY 3034 GATGACGATGCGCACTACACCATCATGCGCAACCCCGAGGGAGAAATCAGCTGT 3090
 Db 457 AspAspAspGlyAsnTyThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 475

RESULT 6

Q8WX93 PRELIMINARY; PRT; 1106 AA.
 AC Q8WX93;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Myonin.
 GN MYN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lockwood S.K., Sims K.B.;

RT "A protein with two immunoglobulin-like domains interacts with the

norrie disease gene product norrin.;"

RL Am. J. Hum. Genet. 61:Al58-Al58(1997).

RN [2]

RP SEQUENCE FROM N.A.

RA Lockwood S.K.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF464873; AAL69964.1; -

DR InterPro; IPR003599; Ig.

DR InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR000634; S/T_dehydrtse.

DR Pfam; PF00047; Ig; 5.

DR SMART; SM00409; IGC2; 5.

DR PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN.1.

SQ SEQUENCE 1106 AA; 122019 MW; 80B52268DD90E7C6 CRC64;

Alignment Scores:

Pred. No.: 1.52e-10 Length: 1106

Score: 19.00 Matches: 19

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.44% Indels: 0

DB: 4 Gaps: 0

US-09-818-990B-1 (1-3963) x Q8WX93 (1-1106)

QY 3034 GATGACGATGCGCACTACACCATCATGCGCAACCCCGAGGGAGAAATCAGCTGT 3090
 Db 844 AspAspAspGlyAsnTyThrIleMetAlaAlaAsnProGlnGlyArgIleSerCys 862

RESULT 7

Q9CWW1 PRELIMINARY; PRT; 379 AA.
 AC Q9CWW1;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 2410003B16Rik protein.
 GN 2410003B16Rik.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;

RA MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aizawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,

RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.;"

RL Nature 409:685-690(2001).
 DR EMBL: AK010350; BAB26871.1; -.
 DR HSSP: P56276; 1TLK.
 DR MGD: MGI:1919583; 2410003B16Rik.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig; 3.
 DR SMART: SM00409; Ig; 3.
 DR SMART: SM00408; IGC2; 3.
 DR SMART: SM00410; IG_like; 1.
 DR Immunoglobulin domain.
 KW SEQUENCE 379 AA; 42157 MW; 42A96BE0AAC88F25 CRC64;

Alignment Scores:

Pred. No.: 2,95e-07 Length: 379
 Score: 16.00 Matches: 16
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.22% Indels: 0
 DB: 11 Gaps: 0

US-09-818-990b-1 (1-3963) x Q9CWW1 (1-379)

QY 3034 GATGACGATGGCAACTACACATCATGCGCAGCCAAACCCCGAGGGAGA 3081
 |||||
 Db 64 AspAspGlyAsnThrIleMetAlaAlaAsnProGlnGlyArg 79

RESULT 8

Q9ET54 PRELIMINARY; PRT; 492 AA.
 AC Q9ET54;
 DT 01-WAR-2001 (TrEMBLrel. 16, Created)
 DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Actin-associated protein palladin (Fragment).
 GN 2410003B16Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A.

RC STRAIN=SWISS WEBSTER/NIH;
 RX MEDLINE=20391984; PubMed=10931874;
 RA Parast M.M., Okey C.A.;
 RT "Characterization of palladin, a novel protein localized to stress
 fibers and cell adhesions.";
 RL J. Cell Biol. 150:643-656(2000).
 DR EMBL: AF205078; AAG00078.1; -.
 DR HSSP: P56276; 1TLK.
 DR MGD: MGI:1919583; 2410003B16Rik.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF00047; Ig; 3.
 DR SMART: SM00409; Ig; 3.
 DR SMART: SM00408; IGC2; 3.
 DR SMART: SM00410; IG_like; 1.
 KW Immunoglobulin domain.
 FT NON_TER 1
 SQ SEQUENCE 492 AA; 55072 MW; BC59E5B3E3BAEBD5 CRC64;

Alignment Scores:

Pred. No.: 2,85e-07 Length: 492
 Score: 16.00 Matches: 16
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.22% Indels: 0
 DB: 11 Gaps: 0

US-09-818-990b-1 (1-3963) x Q9ET54 (1-492)

QY 3034 GATGACGATGGCAACTACACATCATGCGCAGCCAAACCCCGAGGGAGA 3081
 |||||
 Db 177 AspAspGlyAsnThrIleMetAlaAlaAsnProGlnGlyArg 192

RESULT 9

Q9YG65 PRELIMINARY; PRT; 156 AA.
 ID Q9YG65;
 AC Q9YG65;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-WAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE0036.
 GN APE0036.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000058; BAA78945.1; -.
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 156 AA; 16252 MW; C0733DF770584F53 CRC64;

Alignment Scores:

Pred. No.: 11.1 Length: 156
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.70% Indels: 0
 DB: 17 Gaps: 0

US-09-818-990b-1 (1-3963) x Q9YG65 (1-156)

QY 752 GTAGTGTCTCCACGCGCGCTCACTG 726
 |||||
 Db 128 ValValSerProAlaAlaSerLeu 136

RESULT 10

Q8TXM0 PRELIMINARY; PRT; 239 AA.
 ID Q8TXM0;
 AC Q8TXM0;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Predicted phosphate-binding enzyme of the TIM-barrel fold.
 GN MK0642.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyraceae;
 OC Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 and monophyly of archaeal methanogens.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL; AE010357; AA001857.1; -.
 KW Complete proteome.
 SQ SEQUENCE 239 AA; 25052 MW; 179093DD7872137 CRC64;

Alignment Scores:

Pred. No.: 10.4 Length: 239
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.70% Indels: 0
 DB: 17 Gaps: 0

US-09-818-990B-1 (1-3963) x Q8TXM0 (1-239)

QY 2865 CCGGAAGTGTGAGTCTCTGTGCAAA 2839

Db 29 ProgluValLeuGluSerLeuVallys 37

RESULT 11

Q9EV74 ID Q9EV74 PRELIMINARY; PRT; 243 AA.
 AC Q9EV74;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BphS protein.
 GN BPHS
 OS Alcaligenes eutrophus (Ralstonia eutropha).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A5; TRANSPOSON-TN4371;
 RX MEDLINE=97188526; PubMed=9037111;
 RA Merlin C., Springael D., Mergey M., Toussaint A.;
 RT "Organisation of the bph gene cluster of transposon Tn4371, encoding
 enzymes for the degradation of biphenyl and 4-chlorobiphenyl
 compounds."
 RL Mol. Gen. Genet. 253:499-506(1997).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=A5; TRANSPOSON-TN4371;
 RX MEDLINE=20092484; PubMed=10628862;
 RA Mouz S., Merlin C., Springael D., Toussaint A.;
 RT "A GntR-like negative regulator of the biphenyl degradation genes of
 the transposon Tn4371."
 RL Mol. Gen. Genet. 262:790-799(1999).
 DR EMBL; AJ012075; CAC05302.1; -.
 DR InterPro; IPR000524; HTH_GntR.
 DR SMART; SM00345; HTH_GNTR; 1.
 SQ SEQUENCE 243 AA; 27190 MW; E90D8EAF2C2C2F2BD CRC64;

Alignment Scores:

Pred. No.: 10.4 Length: 243
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.88% Indels: 0
 DB: 2 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9EV74 (1-243)

QY 344 CTRACTTCTGCCAGGATAACCCTCGAA 370

Db 111 LeuThrSerAlaArgIleThrLeuGlu 119

RESULT 12

Q9C196 ID Q9C196 PRELIMINARY; PRT; 411 AA.
 AC Q9C196;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CAMP dependent protein kinase regulatory subunit.
 GN PKAR.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ASPERGILLUS NIGER N400;
 RA Panneman H., Rulijter G., Van de Vondervoort P., Legisa M.,
 RT "Cyclic AMP-dependent protein kinase is involved in morphogenesis of
 Aspergillus niger."
 RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AJ296317; CAC36308.1; -.
 DR HSSP; P00515; 2APK.
 DR InterPro; IPR002373; cAMP_kin.
 DR InterPro; IPR000595; cNMP_binding.
 DR Pfam; PF00027; cNMP_binding; 2.
 DR PRINTS; PR00103; CAMPKINASE.
 DR SMART; SM00100; cNMP; 2.
 DR PROSITE; PS00888; cNMP_BINDING_1; 2.
 DR PROSITE; PS00889; cNMP_BINDING_2; 2.
 DR PROSITE; PS50042; cNMP_BINDING_3; 2.
 KW Kinase; cAMP.
 SQ SEQUENCE 411 AA; 44522 MW; BA43E6C393F09706 CRC64;

Alignment Scores:

Pred. No.: 9.7 Length: 411
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.70% Indels: 0
 DB: 3 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9C196 (1-411)

QY 2378 GTVGGTCTGTGGGCTGTGGGAGT 2352

Db 227 ValGlySerValGlyProGlyGlySer 235

RESULT 13

Q9K742 ID Q9K742 PRELIMINARY; PRT; 490 AA.
 AC Q9K742;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein BH3531.
 GN BH3531.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AF001519; BAB07250.1; -.
 DR InterPro; IPR003540; Binary_toxinA.
 DR Pfam; PF03496; Binary_toxA; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 490 AA; 58355 MW; 2258B61391D47319 CRC64;

Alignment Scores:

Pred. No.: 9.47 Length: 490
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.70% Indels: 0
DB: 16 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9K742 (1-490)

QY 202 TCAGAAAGCTGAAAGATCTGGAAGGT 176

Db 398 SerGluargLeuysaspLeuGluGly 406

RESULT 14

O88831 ID O88831 PRELIMINARY; PRT; 587 AA.

AC O88831;

DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ca+/calmodulin-dependent protein kinase beta (Cam-kinase kinase beta).

DE Rattus norvegicus (Rat).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=CEREBELLUM;

RA Fujisawa H.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=CEREBELLUM;

RX MEDLINE=97420710; PubMed=9276695;

RA Kitani T., Okuno S., Fujisawa H.;

RT "Molecular cloning of Ca2+/calmodulin-dependent protein kinase kinase

beta.";

RL J. Biochem. 122:243-250(1997).

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AB018081; BAB3524.1; -.

DR HSSP; Q63450; 1A06.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002290; Ser_thr_pkinase.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00220; S_TKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 587 AA; 64446 MW; C9E49B72578F3971 CRC64;

Alignment Scores:

Pred. No.: 9.23 Length: 587
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 11 Gaps: 0

US-09-818-990B-1 (1-3963) x O88831 (1-587)

QY 2180 TCCCTCCAGACACACCGACGACAA 2206

Db 571 SerProProargThrProProGlnGln 579

RESULT 15

Q9LT87

ID Q9LT87 PRELIMINARY; PRT; 663 AA.

AC Q9LT87;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Similarity to serine/threonine kinase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=20277480; PubMed=10819329;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones.";

RL DNA Res. 7:131-135(2000).

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AB025624; BAB02454.1; -.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002290; Ser_thr_pkinase.

DR InterPro; IPR004040; STY_pkinase.

DR InterPro; IPR001245; Tyr_pkinase.

DR Pfam; PF00069; pkinase; 1.

DR PRINTS; PR00109; TYRKINASE.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00221; STYKC; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 663 AA; 73036 MW; 349C459F290A3C3B CRC64;

Alignment Scores:

Pred. No.: 9.08 Length: 663
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.68% Indels: 0
DB: 10 Gaps: 0

US-09-818-990B-1 (1-3963) x Q9LT87 (1-663)

QY 751 ACACGAGGTCTTCCCTTCATCTCTG 777

Db 211 ThrProGlySerSerProSerSerIeu 219

Search completed: November 30, 2002, 18:48:31

Job time : 144.5 secs

